



OPEN ACCESS

EDITED BY
Leszek A. Majewski,
The University of Manchester,
United Kingdom

REVIEWED BY
Wei Deng,
Soochow University, China
Dongsheng Tang,
Hunan Normal University, China

*CORRESPONDENCE
Nan Du,
nan.du@uni-jena.de

SPECIALTY SECTION
This article was submitted to
Semiconducting Materials and Devices,
a section of the journal
Frontiers in Electronic Materials

RECEIVED 15 August 2022
ACCEPTED 15 September 2022
PUBLISHED 07 October 2022

CITATION
Wang W, Kvatinsky S, Schmidt H and
Du N (2022), Review on data-centric
brain-inspired computing paradigms
exploiting emerging memory devices.
Front. Electron. Mater. 2:1020076.
doi: 10.3389/femat.2022.1020076

COPYRIGHT
© 2022 Wang, Kvatinsky, Schmidt and
Du. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Review on data-centric brain-inspired computing paradigms exploiting emerging memory devices

Wei Wang^{1,2}, Shahar Kvatinsky³, Heidemarie Schmidt^{1,2} and Nan Du^{1,2*}

¹Institute for Solid State Physics, Friedrich Schiller University Jena, Jena, Germany, ²Department of Quantum Detection, Leibniz Institute of Photonic Technology (IPHT), Jena, Germany, ³Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion—Israel Institute of Technology, Haifa, Israel

Biologically-inspired neuromorphic computing paradigms are computational platforms that imitate synaptic and neuronal activities in the human brain to process big data flows in an efficient and cognitive manner. In the past decades, neuromorphic computing has been widely investigated in various application fields such as language translation, image recognition, modeling of phase, and speech recognition, especially in neural networks (NNs) by utilizing emerging nanotechnologies; due to their inherent miniaturization with low power cost, they can alleviate the technical barriers of neuromorphic computing by exploiting traditional silicon technology in practical applications. In this work, we review recent advances in the development of brain-inspired computing (BIC) systems with respect to the perspective of a system designer, from the device technology level and circuit level up to the architecture and system levels. In particular, we sort out the NN architecture determined by the data structures centered on big data flows in application scenarios. Finally, the interactions between the system level with the architecture level and circuit/device level are discussed. Consequently, this review can serve the future development and opportunities of the BIC system design.

KEYWORDS

brain-inspired computing (BIC), memristive computing, artificial synapses, artificial neurons, vector–matrix multiplication (VMM), neural networks (NNs), data structure, system efficiency

1 Introduction

Information processing in the human brain in an analogue and cognitive manner is a key challenge for a brain-inspired computing (BIC) paradigm. The BIC paradigm aims to solve problems using working principles in the brain and drive the next wave of computer engineering (Kendall and Kumar, 2020). The BIC system has been widely used in artificial intelligence (AI) applications such as object detection and classification (Merolla et al., 2014; Pei et al., 2019; Roy et al., 2019), accelerators (Chen et al., 2016; Friedmann et al.,

2016), intelligent robots (Zhang et al., 2016), in-datacenter performance analysis (Jouppi et al., 2017), LASSO optimization problems (Davies et al., 2018), and brain drop (Neckar et al., 2018). All these applications challenge the performance and system efficiency of each module of the BIC system.

In the recent years, based on the extensive research, neural networks (NNs) are considered as efficient methods for the advent of big data, and the proliferation of data and information based on the von Neumann architecture and breakthroughs have been made in terms of improved availability of big data flows, operating power, and training methods (Lawrence et al., 1997; LeCun et al., 2015; Goodfellow et al., 2016). Since 1943, the concept of NNs was first introduced by McCulloch and Pitts (1943), and later, NNs have been widely studied and developed. In 1957, a machine that simulated human perception was proposed by Frank Rosenblatt, which is known as single-layer perceptron (SLP). The SLP was the first generation of NNs, with single-feature layers that could be applied to recognize some letters of the alphabet. In 1985, the multiple hidden layers were applied in the perceptron to replace the only single layer by Geoffrey Hinton, which is known as multilayer perceptron (MLP), and this was the beginning of the second generation of NNs. As the range of applications for NNs expands, various NN structures have since emerged, such as the convolutional neural network (CNN), graph neural network (GNN), recurrent neural network (RNN), and their various variants. Despite the existence of numerous types of NNs, there is still a fundamental challenge to be able to realistically and accurately emulate the human brain. Thus, the third-generation NNs represented by the spiking neural network (SNN) have emerged, which is considered to be the closest to the synapses and neurons of the human brain.

While exploring the advanced BIC paradigms, some reviews have been carried out in terms of applied materials (Ko et al., 2020; Liu et al., 2020) or devices (Kwon et al., 2022; Meier and Selbach, 2022) or neuromorphic computing algorithms (Kumar et al., 2022; Yang et al., 2022). The two-dimensional (2D) materials were reviewed to apply for the next-generation computing technologies (Liu et al., 2020), which is focusing on a broader application scope in neuromorphic computing, matrix computing, and logic computing. The review work on emerging neuromorphic devices by using 2D materials (Ko et al., 2020) comprehensively explores various 2D materials, especially their prospects for neuromorphic applications. Furthermore, in the recent years, the memristors, owing to their non-volatile and reconfigurable properties, are considered as the promising candidates for BIC systems. The 2D memristive devices were reviewed for the applications of neuromorphic computing, and the fabrication and characterization of neuromorphic memristors were primarily discussed (Kwon et al., 2022). Moreover, the research progress on memristors especially as artificial synapses to neuromorphic systems is reviewed by Yang et al. (2022).

In comparison to aforementioned reviews, in this work, we overview the recent development of BIC systems, with respect to the perspective of a system designer. We provide a comprehensive review on the advances in the device level and circuit level up to the architecture and system levels for constructing a reliable BIC system, especially pointing out the interaction among them. Particularly, in the data-centric era, the structures of data sets in diverse application scenarios have a significant influence on the construction of NNs. Hence, the choice of NN algorithms sorted by data structures is emphasized in this work. The work is structured as follows: the functional materials and devices for BIC systems in terms of low-dimensions including zero-dimension (0D), one-dimension (1D), and two-dimensions (2D) are reviewed in Section 2. The artificial synapses and neurons as the building blocks of the BIC system are discussed in Section 3. NN algorithms and architectures, which are determined by the data structures in AI applications, are introduced and discussed in Section 4. Section 5 discusses the interactions among the device, circuit, and architecture from system level aspects, which include the pros and cons among the 0D, 1D, and 2D materials in the device level, the features of artificial synapses and neurons based on memristors in the circuit level, various neural networks depending on data structures in the architecture level, and the current challenges and perspectives at the device/circuit level, architecture level, and system level, which provide guidance for future research.

2 Materials and devices for brain-inspired computing systems

The device level is crucial for the hardware of the BIC system. Based on different functional materials, diverse devices can be realized for constructing circuits in the BIC system. Functional materials exist in many ways according to different functional requirements and classifications. For example, the main difference between organic and inorganic materials is that organic compounds always contain carbon, while most inorganic compounds do not. The organics [i.e., polymers (Biesheuvel et al., 2011; Fong et al., 2017; Namsheer and Rout, 2021), organic molecular crystal (Wang et al., 2011; Lee et al., 2012) and carbon allotropes (Zestos et al., 2014; Schwerdt et al., 2018; Xu et al., 2021), and hydrogel composites (Shi et al., 2014; Kayser and Lipomi, 2019; Pyarasani et al., 2019)], owing to unique features such as long-term biocompatibility, good mechanical flexibility, and molecular diversity, have been explored in neuromorphic devices (Deng et al., 2019; Tuchman et al., 2020; Go et al., 2022). In addition, various inorganics with unique advantages are not being ignored (i.e. metal oxides (Hu et al., 2018; Gao et al., 2022), sulfur compounds (Knoll et al., 2013; Lu and Seabaugh, 2014; Wu et al., 2017), and halogenated compounds (Cheng et al., 2021)) in many

TABLE 1 Summary of the reported 0D, 1D, and 2D devices for neuromorphic computing.

Type	Material	Device	Endurance (cycle #)	Operational current (μA)	Operational voltage (V)	Retention (s)	Number of states	Reference
0D devices	Quantum dots	FTO/MoS ₂ -QDs/Al	60	~16	± 2	10 ³	N/A	Thomas et al. (2020)
		RGO/NCQD/graphene	N/A	~25	± 2	10 ³	N/A	Lin et al. (2020)
	Nano particles	Ag/CiC:Ag/ITO	10 ⁶	~90	1/-0.4	10 ⁴	250	Ge et al. (2020)
1D devices	Nanowires	Pt/ZnO nanowires/Ag	3,000	10	5/-1.5	>10 ³	N/A	Milano et al. (2018a)
	Block polymer	Au/block copolymer polyacrylamide/poly (acrylic acid)/Au	N/A	0.01	2.9	N/A	N/A	Zhitenev et al. (2007)
	CNT	Ti/CNT/Al	N/A	$\sim 5.2 \times 10^{-3}$	± 6	N/A	N/A	Kim et al. (2013)
2D devices	TMCs	WSe ₂	N/A	~0.001	1.2/-0.4	>5 $\times 10^3$	60	Zhu et al. (2018)
		MoS ₂	>10 ³	~0.01	± 4	>7 $\times 10^3$	100	Zhu et al. (2019)
		MoS ₂	>10 ³	~10 ⁻⁴	5/-4	>10 ⁴	100	Wang H. et al. (2020)
	h-BN	h-BN/WSe ₂	N/A	~10 ⁻³	± 0.3	N/A	600	Seo et al. (2018)
		Ti/h-BN/Au	500	~0.01	0.8/-0.7	10 ²	N/A	Shi et al. (2018)
	Graphene	Multi-layer graphene/Dy ₂ O ₃ /indium tin oxide (ITO)	>10 ²	20	± 0.2	>10 ⁴	N/A	Zhao et al. (2014)
	TaO _{2-x}	Pt/A ₂ O _{5-x} TaO _{2-x} /Pt	10 ¹²	10/30	2/-1	10 years	N/A	Lee et al. (2011)

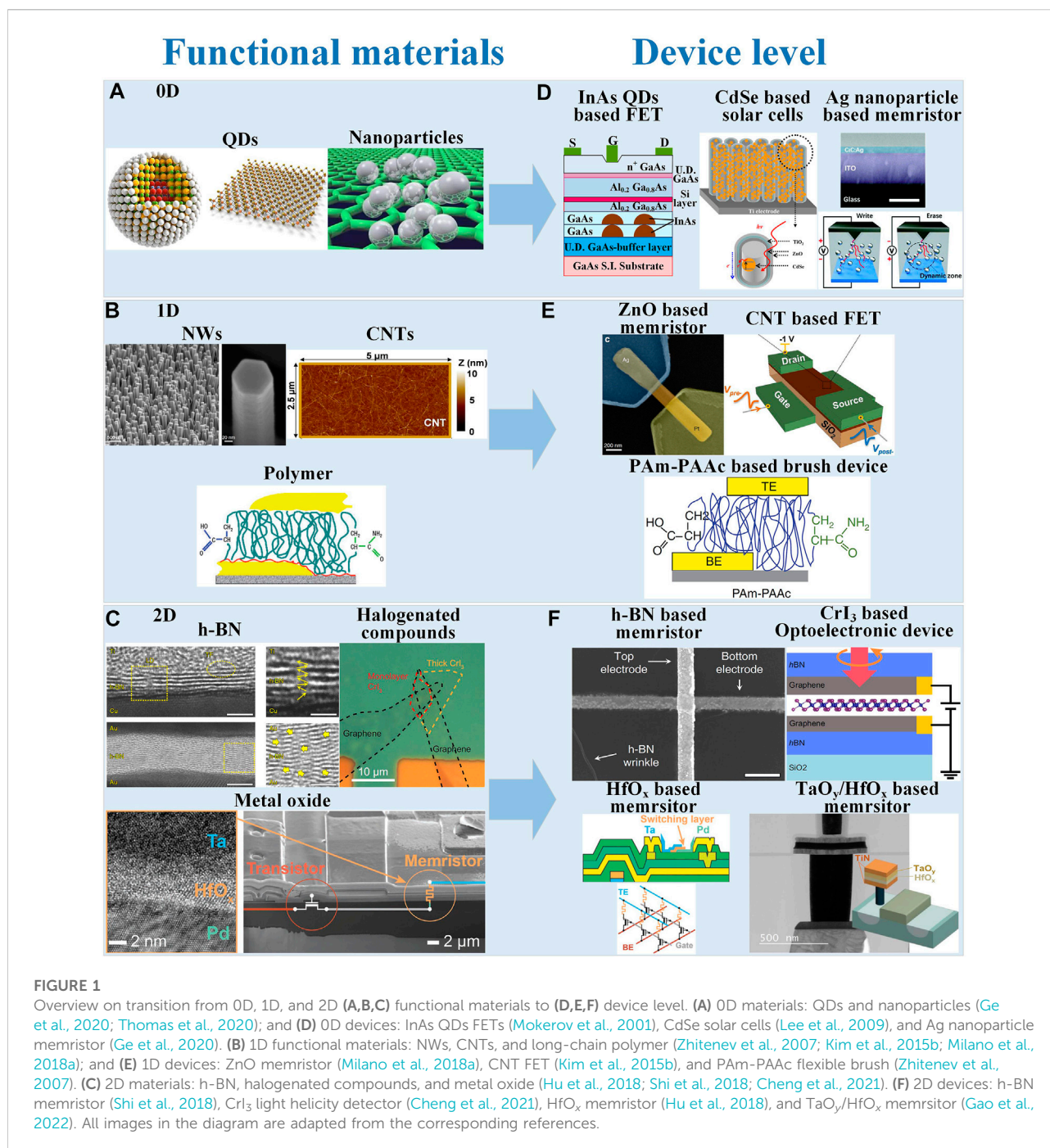
applications. Another way is according to different dimensional orientation; the materials can be classified as 0D, 1D, and 2D. It can be referred as 0D material when it is within a nanoscale range of (1–100 nm) in three dimensions or if they are composed of basic units. 1D materials refer to those in which electrons are free to move in only one nanoscale direction (linear motion), such as nanowire (NW) junction materials, quantum wires, most representative carbon nanotubes (CNTs), and polymers. 2D materials refer to materials in which electrons are free to move in a planar between 1–100 nm, such as graphene, h-BN, metal oxides, sulfur compounds, and halogenated compounds. In this section, we discuss the functional materials and devices for the BIC system in terms of low-dimensions including 0D, 1D, and 2D. Table 1 summarizes the reported 0D, 1D, and 2D devices for neuromorphic computing.

2.1 Zero-dimensional devices

0D devices are composed of the materials based on 0D structure, such as semiconducting quantum dots (QDs) and nanoparticles, as demonstrated in Figure 1A. First, 0D materials are suitable for application in neural morphology-related photonic systems, owing to its promising optical properties. Attributing to the fact that 0D photons are not limited in space and power density during propagation, 0D

photonic devices are used for parallel communication and super-connectivity. In the context of neuromorphic computing, 0D materials show great potential in many applications, such as Ag/Au nanoparticles (Alibart et al., 2010; Ge et al., 2020), InAs/InGaAs-based QDs (Mesaritakis et al., 2016), black phosphorus-based QDs (Han et al., 2017), MoS₂-based QDs (Thomas et al., 2020), GaAs-based QDs (Lee et al., 2009), and CdSe-based QDs (Moreels et al., 2007) as listed in Figures 1A,D and Table 1. The local gain can be obtained by size-tunable plasmonic responses from 0D metal nanoparticles. Attributing to charge-trapping actions, 0D Au nanoparticles have been exploited in synaptic transistors (Alibart et al., 2010). The phenomena of dynamic contraction and extension can be seen with these Au nanoparticles, in which the competing effects between surface tension and electric field play a role, which leads to effective learning behavior. In addition, the synergistic effect between Ag nanoparticles and electrodes can be applied for flexible neuromorphic networks to enhance switching properties (Ge et al., 2020) as listed in Table 1.

The excitatory and inhibitory synaptic responses have been imitated by the InAs/InGaAs-based QDs with explicit energy levels, in which multi-band emission from QDs enables operation (Mesaritakis et al., 2016). But a challenge for integrating QDs and photonic waveguides still remains. The site-controlled QDs were employed to realize electro-photo-sensitive devices on GaAs/AlGaAs wafers with tuning



conductance by electrical/optical pulses (Maier et al., 2015; Maier et al., 2016). In addition, the metal and QD nanoparticles were embedded in a matrix manner to obtain resistive random access memory (ReRAM or RRAM). In particular, the black phosphorus QDs are sandwiched within polymers (i.e., methyl methacrylate) with bilayers to print multi-level ReRAM, presenting high switching ratio up to $\sim 10^7$, in which the filament formation and rupture are achieved by the charge

trapping being similar with Au nanoparticles in QDs (Alibart et al., 2010). Therefore, the quantum memristor devices can also obtain opportunities from QDs, especially for Josephson junction-based devices, in which a state variable can be the phase difference between quasi-particles (Chua, 2003; Pershin and Di Ventra, 2011). MoS₂ is valuable in QD devices, attributing to the bob radius of 23 nm, which is very favorable for achieving size quantization. Thus, after being quantized, MoS₂-based QDs

can then be applied to 0D devices, such as ReRAM, and quantitative MoS₂ can act as top/bottom electrodes in the QDs neuromorphic devices (Thomas et al., 2020) as listed in Table 1. This quantitative manipulation of the material can significantly increase its utilization and expand the range of applications. In addition, QDs can be used for the realization of targeted nanostructures, where the electrons in the specific structure are confined to exhibit specific properties (Mukherjee et al., 2016). Therefore, it is also used to modulate doped FETs (MODFETs) as a means of changing their I–V characteristics. It was found that by embedding InAs-based QDs in GaAs channels (Figure 1D), MODFETs exhibit high field I–V characteristics attributed to the electrons in the QDs, and compared to conventional MODFETs, QD devices exhibited hot-electron transistors that would be valuable in high-speed applications (Mokerov et al., 2001). Similarly, CdSe-based QDs (Figure 1D) are used in solar cell devices, owing to high light absorption properties in the visible region, replacing the previous Ru sensitizer, attributing to QDs a higher extinction coefficient with $\sim 10^5$ and quantum confinement effect (Lee et al., 2009). It is undeniable that these 0D devices can overcome the traditional defects and improve the switching ratio and speed of circuits and even systems. The quantum NNs based on QD array are also proposed. In addition, the electrochemical neuromorphic organic device presents low energy (< 10 pJ for $10^3 \mu\text{m}^2$ devices), and non-volatile conductance states are more than 500 within the range of ~ 1 V, and high accuracy can be acquired in NN applications as listed in Table 1.

2.2 One-dimensional devices

1D devices are composed of the materials based on the 1D structure, such as nanowires (NWs), carbon nanotubes (CNTs), and polymers. Early on, 1D materials were extensively studied, showing that they share a very similar topology with tubular axon, which is the key to achieving hyperlinkability in the biological system. As 1D materials are being studied widely, their diverse properties in both physics and chemistry, solution processability, and the bottom-up growth feature compared to 0D materials would allow 1D materials to behave more potentially in BIC systems.

1D semiconductor NWs have been developed worldwide as low-dimensional single crystals. 1D metal oxide NWs fabricated using a bottom-up approach are being investigated for the implementation of resistive switching devices as this method can achieve device size reduction beyond limitations of traditional lithography (top-down approach) and can be considered a good platform for highly localized and characterized switching events (Nagashima et al., 2011; Ielmini et al., 2013; Milano et al., 2018a) as listed in Figures 1B,E and Table 1. For such reasons, such studies have been carried out not only in NWs (Oka et al., 2009; Nagashima

et al., 2010; He et al., 2011; Nagashima et al., 2011; Yang et al., 2011; Ielmini et al., 2013; Qi et al., 2013; Liang et al., 2014; O'Kelly et al., 2014), but also in arrays of NWs (Park et al., 2013; Anoop et al., 2017; Porro et al., 2017; Xiao et al., 2017; Milano et al., 2018b), which have become another major material for 1D devices. However, a non-negligible problem regarding NWs is the tendency of Joule heating to induce them to melt and lead to further hardware failures; thus, single devices based on NWs still suffer from high operating voltages or poor device reliability in terms of ruggedness and variability (Oka et al., 2009; Nagashima et al., 2010; He et al., 2011; Nagashima et al., 2011; Yang et al., 2011; Ielmini et al., 2013; Qi et al., 2013; Liang et al., 2014; O'Kelly et al., 2014).

The CNT shows many unique properties in addition to those similar to NWs as listed in Figures 1B,D. It is a graphite cylinder, which shows semiconductor/metallic behaviors according to its chiral vector and has been used to construct 1D devices. Single-walled CNTs have been used in post-silicon digital logic because of their high charge carrier mobility and scalable properties, especially in the ultra-short channel limit (i.e., sub-5 nm nodes) (Cao et al., 2017; Milano et al., 2018b). CNTs have been used in transistor devices and as models for synaptic circuits (Joshi et al., 2009; Kim et al., 2015b). However, due to the challenges faced by individual CNTs in practical applications, such as wafer-level assembly of CNTs and alignment aspects, efforts are applied to study CNT networks structured as thin-film transistors (TFTs) (Milano et al., 2018b). CNTs can be used widely as aligned arrays (Sanchez Esqueda et al., 2018) and random networks (Kim et al., 2013; Shen et al., 2013; Kim et al., 2015b; Feng et al., 2017; Kim et al., 2017; Danesh et al., 2019) as listed in Table 1, which have also indirectly demonstrated the research potential and value of synaptic transistors. In addition, these CNT-based 1D devices have also been applied to neuromorphic circuits for unsupervised learning in NNs (Sanchez Esqueda et al., 2018). However, the lack of NVM and the inherent problem of transverse geometry in CNT-based devices results in the inability to form dense arrays of horizontal bars.

Furthermore, the hybridization like the ZnO nanowires (1D) decorated with CeO₂-based QDs (0D) were applied in memory devices (Younis et al., 2013), which were investigated for charge-trapping mechanisms with the aim of designing more advanced devices. Similarly, the diffusion of Ag along the ZnO nanowires after they are in contact is observed (Figure 1D). The diffusion behavior can produce volatile/non-volatile resistive switching, which is similar to Ag-SiO_x diffusion-based memristor devices (Wang et al., 2017; Milano et al., 2018a).

In particular, the polymers not only have the characteristics and research value of organic materials in neuromorphic devices to boost the properties of devices and broaden the application areas (c), but they also possess 1D properties. Therefore, polymers also possess 1D properties that have been exploited in 1D devices such as memories, organic electrochemical transistors (OECTs), and nanoscale molecular devices

(Zhitenev et al., 2007; Waser et al., 2009; Cho et al., 2011; Rivnay et al., 2018; van De Burgt et al., 2018) as listed in Figures 1B,D and Table 1. The synaptic transistors can be enabled by combining polymer electrolytes with planar silicon and lithium-ion battery materials (Lai et al., 2010; Fuller et al., 2017). In addition, OECTs show promising conductivity states (> 500) within 1 V range and low energy consumption (~ 10 pJ per event) (Van De Burgt et al., 2017; Fuller et al., 2019).

2.3 Two-dimensional devices

2D devices are composed of the materials based on the 2D structure, such as graphene, MoS₂, TMCs including MX₂ (M = Mo, W; X = S, Se), and hexagonal boron nitride (h-BN) (Figures 1C,F and Table 1). The widespread use of 2D materials is due to three unique features: First, the layers in 2D materials are all bonded by covalent bonds, and the van der Waals forces between adjacent layers are very small, so that layer-by-layer exfoliation of 2D bulk materials can be achieved. Second, in the exfoliated ultrathin 2D bulk structure, the electron activity space is limited, so it can be precisely controlled by the gate voltage to eliminate the influence of the short channel effect. Third, the 2D material system is relatively large, and innovation can be achieved by controlling the energy band structure. In addition, 2D devices show scaling and integration with planar wafer technology. The development of new 2D device concepts is spurred subsequently by realizing neuromorphic functionality in 2D nanomaterials to reveal unexpected mechanisms. Although 2D devices have shown great advantages and value, they are difficult to prepare in large quantities, have low preparation efficiency, are prone to defects, are prone to the introduction of impurities, are difficult to control the composition of the product and have high environmental requirements, and preventing large-scale industrialization. This is why materials with mature preparation processes such as graphene and h-BN have endured for so long.

2D nanomaterials attracted researchers as new building blocks for the development of memristor-based (Jin et al., 2015; Lee et al., 2016; Lei et al., 2016; Schmidt et al., 2016; Vu et al., 2016; Sangwan and Hersam, 2018; Wang T.-Y. et al., 2020; Chen et al., 2021) and non-memristor-based devices (Geim and Grigorieva, 2013; Jariwala et al., 2014), which are essential parts in the BIC system. Here, we refer to 2D memristors-based devices (Bessonov et al., 2015; Cheng et al., 2016; Tian et al., 2017; Huh et al., 2018; Sangwan et al., 2018; Shi et al., 2018; Jadwiszczak et al., 2019; Wang L. et al., 2019; Zhang et al., 2019; Zhong et al., 2020). A typical metal-insulator-metal (MIM) is the structure of memristor and also includes a 2D insulator as the switching layer. The reason why 2D materials can be used for artificial synaptic devices in the field of BIC is attributed to its unique structure, electronic properties, and mechanical properties (Arnold et al., 2017; Bao et al., 2019). The Ta/HfO_x/Pd memristor is applied to

accelerate computations in the BIC system (Hu et al., 2018) as listed in Figure 1F. MoS₂, as one of promising 2D materials, is used in a wide range of applications due to its attractive direct band gap structure, which creates its large conductivity and electron mobility. Following this, TMCs, including transition metal dichalcogenides (TMDs) (Kwon et al., 2022), have become a research hotspot (Fiori et al., 2014). The ultrathin 2D memristor devices (thickness < 1 nm) based on TMCs demonstrate high-frequency switches (Lu and Seabaugh, 2014; Wu et al., 2017), due to the low ON-state resistance (< 10 Ω). The conventional thinking, which asserts the leakage currents still exist in monolayer 2D semiconductors, has been refuted by their high switching ratio ($> 10^4$) (Wu et al., 2017). The results indicate the emergence of new switching mechanisms, in which point defects may be a key (Lu and Seabaugh, 2014). In addition, the memristor devices with the switching layer of few-layer MoS₂ and electrode layers of graphene demonstrate high operating temperature (340°C) than conventional memristor devices with metal oxide (200°C) (Kim et al., 2020). The bilayer MoS₂ 2D memristor devices with electrode of Cu and Ag present low switching voltages (~ 0.2 V) (Knoll et al., 2013).

By taking full advantage of the respective fascinating properties of 2D materials, 2D devices performances can be boosted. As one of the first 2D materials to be discovered, graphene is used as a switching layer in devices to improve their stability. In addition, it is used as a contact material to boost robust features for memory cells (Jariwala et al., 2014). As a material with a structure very similar to graphene, h-BN, also known as white graphene, exhibits properties distinct from those of graphene, such as high insulating properties, which are different from graphene (Sangwan and Hersam, 2018). It is employed as a switching layer in devices, endowing devices with high endurance and low operating currents (Shi et al., 2018) as listed in Table 1. 2D memory devices are able to operate at low operating voltages and therefore have low energy consumption, which is attributed to one of the promising features of 2D materials, namely, the ultra-thin structure (Cheng et al., 2016; Zhang et al., 2019) as listed in Table 1. Consequently, the matrix computing can be boosted by potential 2D memory devices with a promising structure in operating voltage/current, accuracy, etc., which is also a frequent concern for devices.

One of the most promising performances in matrix computing is extremely low energy consumption. As can be seen in Table 1, the different materials form devices with different levels of operating voltage and current. Generally, the low operating voltage and ultralow current also mean low power consumption. Recently, some significant efforts have been made for reducing the energy consumption, operating current, and voltage; for example, a memristor crossbar array based on 2D hafnium diselenide (HfSe₂) has been fabricated using the molecular beam epitaxy technique, which exhibited small switching voltage with 0.6 V, especially low switching energy

TABLE 2 2D memristive devices applied as artificial neurons and artificial synapses in the BIC system.

Device	Working principle	Structure (2D switching layer)	Set voltage/current	Reference
Synapse	Conductive filament formation	Cu/MoS ₂ /Au	0.1–0.2 V	Xu et al. (2019)
		Ag/ZrO ₂ /WS ₂ /Pt	0.4 V	Yan et al. (2019)
		Ni/Au/MoS ₂ /Gr	3 V	Ko et al. (2020)
	Vacancy migration	Ni/Au (top), ML-MoS ₂ /1L graphene, and graphene (bottom)	1.5 V	Krishnaprasad et al. (2019)
		Pd/WS ₂ /Pt	0.56–0.67 V	Yan et al. (2019)
		Ag/WO _{3-x} /WSe ₂ /Gr	3 V	Huh et al. (2018)
		Ag/MoO _x /MoS ₂ /Ag	0.15 V	Bessonov et al. (2015)
	Phase transition	Gr/MoS _{2-x} O _x /Gr	3.5 V	Wang et al. (2018)
Au/MoS ₂ /Au		4 V	Zhu et al. (2019)	
Neuron	Conductive filament formation	Ag/MoS ₂ /Ag	0.35–0.4	Dev et al. (2020)
		Ag/MoS ₂ /TiW	1.2V/1 μA	Hao et al. (2020)
		Cu/Mxene/Cu	1V/0.1 μA	Chen et al. (2019)
	Grain boundary mediated transport	Ni (top), ML-MoS ₂ /1L graphene, graphene (bottom)	2.8–2.9 V	Kalita et al. (2019)

with 0.82 pJ (Li et al., 2022). In addition, a memristor device was prepared with h-BN materials through chemical vapor deposition by Shi et al. (2018) as listed in Figures 1C,F. The device presented excellent performances with ultralow power consumption of 0.1 fW, which cannot be obtained without the role of h-BN. Based on few-layer h-BN as a switching layer, the bipolar/unipolar resistive switching can be achieved in some 2D devices (Gandhi et al., 2011; Ganjipour et al., 2012), in which switching is controlled by controlling the generation of defects from active Cu/Ag. The MoS₂/h-BN/graphene heterostructures were applied in random access memory devices by Vu et al. (2016), which demonstrate ultrahigh on/off ratio of 10⁹, low operating voltage with 6 V, and ultralow off-state current with 10⁻¹⁴.

As aforementioned, the functional 0D, 1D, and 2D materials play a crucial role in constructing devices for BIC systems. In this review, we focus on memristors due to their promising functional properties for constructing BIC systems in the circuit level, architecture level, and system level.

3 Circuit level: Artificial synapses and neurons

In the biological human brain, there are 86 billion neurons that perform activation functions and transmit information *via* around 1,000 trillion synapses, which form the learning functions, that is, senses such as smelling and listening, in the human brain. External signals are received and transmitted to synapses, which in turn are transmitted to neurons. The activation functions are processed to produce new signals, which are then transmitted back to the senses in reverse,

producing the so-called “limbic response.” In this section, we review artificial synapses and neurons in the BIC system. The characteristics of artificial synapse and neuron devices based on 2D memristors are summarized in Table 2, such as set voltage/current, memristor structure, functional materials, and their working principle.

3.1 Artificial synapses

The weight storage and induction protocol unit constitute a simple synapse, which modifies the weights according to the pulse signals. In general, the proliferation of plasticity-related proteins (PRPs) is the main cause of synaptic interactions during which these proteins are required for synaptic growth between adjacent synapses (Farris and Dudek, 2015) as listed in Figure 2A. The synaptic interactions are fundamental and essential in the neuromorphic systems. In the past, two approaches were implemented, one through a software-assisted circuit concept (Sheridan et al., 2017) and the second through a voltage splitter effect assisted by an external voltage bias (Borghetti et al., 2010). However, these two approaches do not directly mimic the internal processes of biological synapses. The ionic coupling involves the diffusion and exchange of ions between artificial synaptic devices that can potentially mimic this process. Thus, the competition and cooperation effects of biological synapses can be enabled appropriately.

The earliest artificial synapse was achieved with transistors (which can be called synaptic transistors) by using a floating gate. The charge of the floating gate (storage or dissipation) can be controlled by the injection and tunneling from hot electrons, respectively. Therefore, the threshold voltage and conductivity of

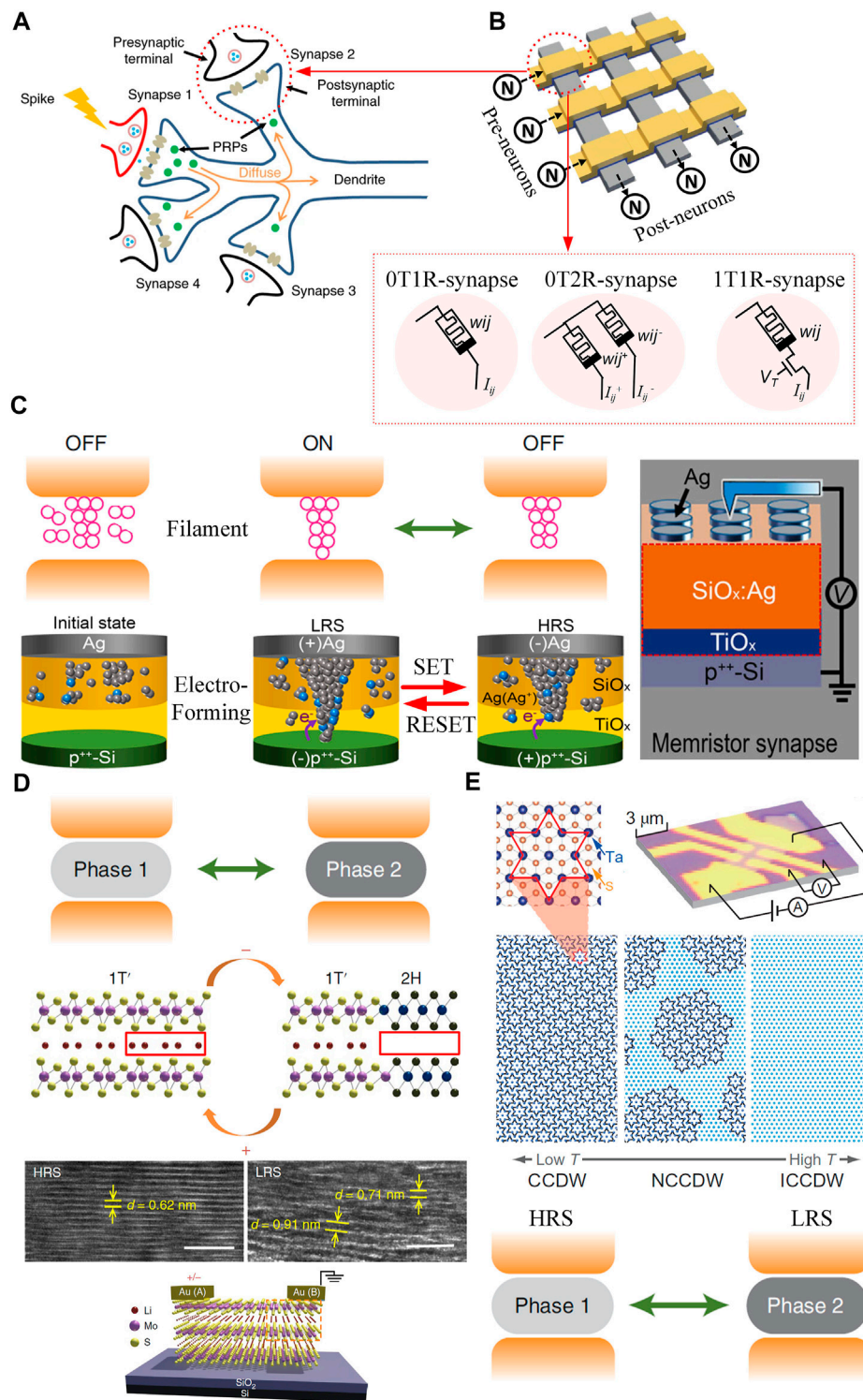


FIGURE 2

Working principle of biological synapses and representative memristive artificial synapses. (A) Illustration of functional connection for artificial neurons and synapses by using MCAs. (B) Configurations of memristive artificial synapses, including 0T1R-, 0T2R-, and 1T1R-synapses (Du et al., 2021a). Schematics of I–V characteristics of different configurations of memristive artificial synapses (Sangwan and Hersam, 2020). (C) Illustration of artificial memristive synapse induced by filament formation and rupture. The inset demonstrates the filamentary switching of a single memristor unit formed by the Ag/SiO_x:Ag/TiO_x/p⁺⁺-Si structure (Zhu et al., 2019; Ilyas et al., 2020; Sangwan and Hersam, 2020). The artificial memristive synapse whose transconductance is controlled by its structural phase (D,E). (D) Insets demonstrate local transition from the 1T' phase (LRS) to the 2H phase (HRS) in Li⁺-MoS₂, which is related with Li⁺ ion migration and controlled by a voltage bias (Zhu et al., 2019; Sangwan and Hersam, 2020). (E) Insets demonstrate phase change by controlling temperature, the CCDW phase at low temperature (left), in the hexagonal NCCDW (middle), and in the ICCDW phase at high temperature (right) (Yoshida et al., 2015). All images in the diagram are adapted from the corresponding references.

the transistor are modulated efficiently (Diorio et al., 1996). For synaptic transistors, the robustness and open channel are critical advantages, which can lead to spatiotemporal responses that are achieved by multiple gate terminals (Buonomano and Maass, 2009; Qian et al., 2017). However, it cannot be ignored that the relatively large footprint is accompanied by lateral geometry from synaptic transistors, which is suboptimal for obtaining synapses with high density (Lanza et al., 2019).

Subsequently, the memristors as a popular emerging device (Figure 2B top), which possess a smaller size (Yu and Chen, 2016; Pi et al., 2019), enable 3D stacked memory (Seok et al., 2014), and combine both an induction protocol and memory into a single device (Jeong et al., 2016), become a suitable alternative in terms of the application of artificial synapses (as listed in Table 2). The functional materials and operating mechanisms used in memristive devices have been reviewed extensively (Waser et al., 2009; Pershin and Di Ventra, 2011; Kuzum et al., 2013; Yang et al., 2013; Tan et al., 2015).

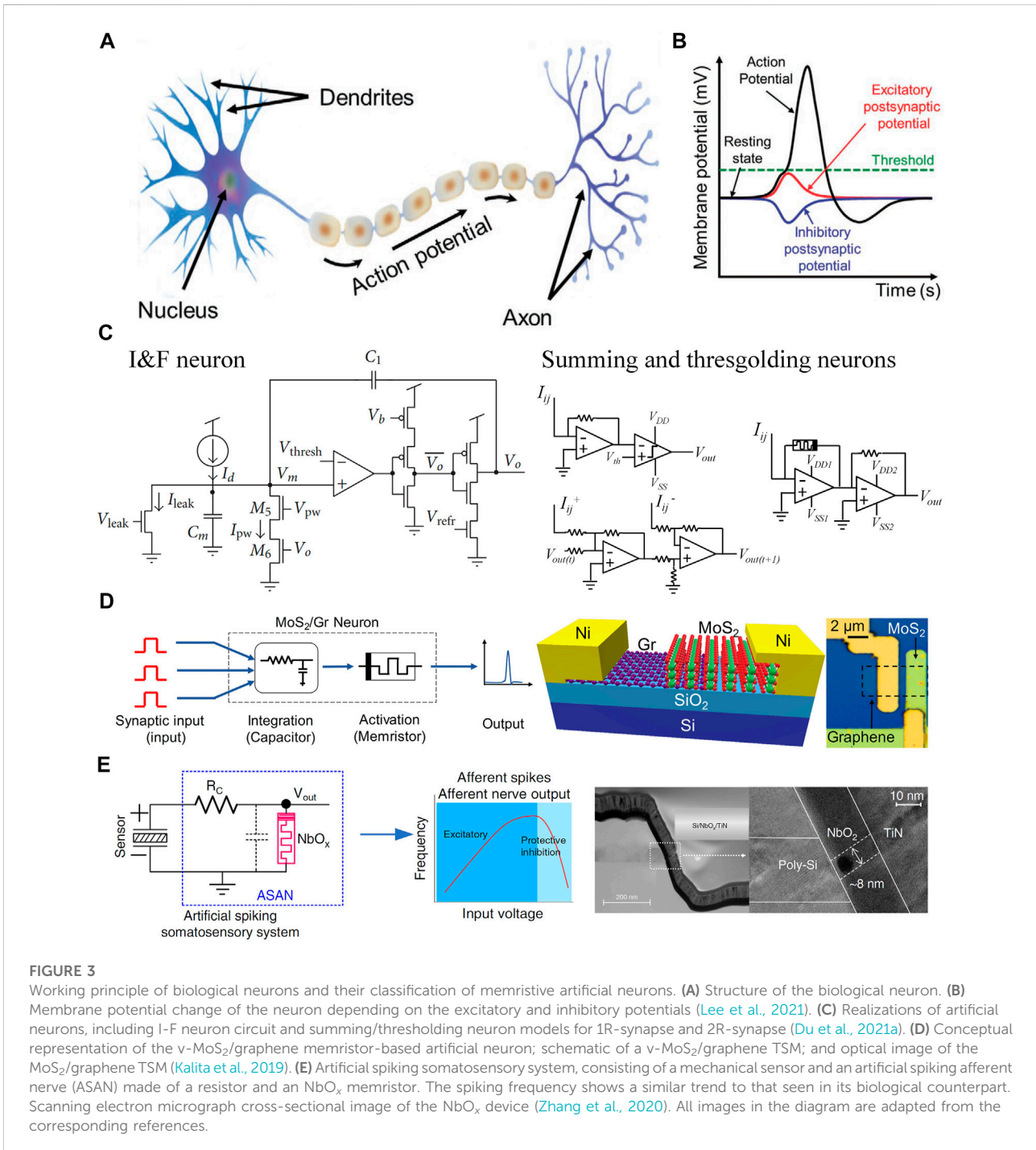
The different configurations of memristor-based artificial synapses are listed at the bottom of Figure 2B. Memristors with 0T1R configuration (T is the transistor, and R is the memristor) have been applied to implement artificial synapses (Krestinskaya et al., 2017; Zhang Y. et al., 2017; Du et al., 2021b), and such synaptic devices are very efficient in terms of density and power consumption. In addition, the non-volatile resistive switching will occur in the memristors with distinct resistance states with zero bias (Yang et al., 2008); however, the common leaky path problems of memristive crossbar arrays (MCAs) cannot be ignored. A transistor in series with the memristors might be a good method for this (Yao et al., 2017), but it is not as good as a 0T1R configuration in terms of density. Another popular configuration for implementing artificial synapses consist of two memristors (i.e., 0T2R) (Alibart et al., 2013; Hasan and Taha, 2014), which have the advantage of being able to double the area size and to apply the negative synaptic weights required in an NN. Voltage stimulation signals generated from neurons (word-lines) and the current signal can be generated from each bit line. In unipolar switching, only at the same bias polarity, the switching events can occur, whereas the devices can be operated to on- and off-states by reverse bias in the bipolar resistive switching (Yang et al., 2008; Chang et al., 2009). Therefore, by connecting the two bipolar switches back-to-back, this results in a complementary resistive switch (Linn et al., 2010).

The 1T1R is the third configuration besides 0T1R and 0T2R, which has been studied for artificial neurons and synapses (Figure 2B bottom). The chips with 1T1R configurations have been widely used (James et al., 2017; Schuman et al., 2017; Krestinskaya et al., 2019), and almost all of them employed crossbar arrays to address individual synapse nodes. In 1T1R configuration, the leakage current can be reduced in the 1T1R configuration, which is attributed to the use of the transistor as a selector. Two memristive elements are included in the

memristor, in addition to the pinched hysteresis loop, which can be modulated by the gate terminals (Mouttet, 2010; Sangwan et al., 2015; Sangwan et al., 2018; Wang L. et al., 2019; Yang et al., 2019). Memristor devices are often complex in both structural design and preparation, but this is also the main reason for the simplicity of the circuit when they are applied to circuits (Chua and Kang, 1976; Pershin and Di Ventra, 2011; Xia et al., 2011; Abdelouahab et al., 2014; Kim et al., 2015a). Thus, for 1T1R configuration, not only do transistors and memristors combine their respective characteristics into one device, but they also exhibit unique features in terms of applications such as spike-timing-dependent plasticity (STDP) and bio-realistic hyper-connectivity (Bi and Poo, 1998; Caporale and Dan, 2008; Sangwan et al., 2018). Precisely, the conversion mechanism of the memristor corresponds well to biological synapses, which can be generated by the formation and rupture of filaments (Figure 2C up) or by modulation of the Schottky barrier by defects or migration of ionic species.

The Ag/SiO_x:Ag/TiO_x/p⁺⁺-Si memristor devices have been reported by Ilyas et al. (2020), and the devices present analog switching behaviors. The schematic diagram of the device is shown in Figure 2C. In such a simple device, the SiO_x:Ag and TiO_x thin layers serve as the transition layer, and the Ag TE and p⁺⁺-Si BE serve as the electrodes, respectively. The Ag/SiO_x:Ag serves as the presynaptic membrane, and the TiO_x/p⁺⁺-Si as the postsynaptic membrane in the Ag/SiO_x:Ag/TiO_x/p⁺⁺-Si memory device. The ions from the synaptic weights are released between the presynaptic and postsynaptic membranes, thus causing a change in synaptic weights when it receives a neural pulse. The Ag ions migrate in response to the voltage pulse, thus causing the conductivity of the Ag/SiO_x:Ag/TiO_x/p⁺⁺-Si memory device to be modulated. The physical model is shown in Figure 2C, and the switching mechanisms of the SiO_x:Ag- and SiO_x:Ag/TiO_x-based memristor devices are similarly explained.

Phase-change memristor devices show a metal-to-insulator transition by local heating and rapid quenching (Figure 2D). The schematic illustration describing the phase transition is shown in Figure 2D, in which electrode A can drive migration of Li⁺ ions. The 1T' phase will produce as the increase of Li⁺ concentration; conversely, the 2H phase can be brought forth by the decreasing of Li⁺ concentration. The high-resolution transmission electron microscopy images demonstrate Li_xMoS₂ film have been switched to the high-resistance state (HRS) and low-resistance state (LRS), respectively, as listed in Figure 2D. The MoS₂ lattice fringes can be observed with uniform interlayer spacing (~0.62 nm) in the HRS sample. However, the distorted MoS₂ lattice fringes with interlayer spacing (for example, 0.91 and 0.71 nm at the two positions) (Zhu et al., 2019) can be observed in Figure 2D. The phase transition of molybdenum ditelluride (MoTe₂) was deployed by Wang et al. (2019b), and another example of polymorphic TMD (Cho et al., 2015) demonstrates that the phase change resistive memory devices were devised. The phase transitions for MoTe₂ occurred by the electric field, in



which the electronic characteristics can be found between both phases.

Figure 2E illustrates schematic illustration of quantum phase transition memristors. Yoshida et al. (2015) show that polymorphic memory non-volatile switching consisting of 1T-TaS₂ with first-order charge density wave (CDW) phase transitions (Sipos et al., 2008; Stojchevska et al., 2014), which

presents the first-order CDW phase transition (Figure 2E). It can be seen that the phase switching occurred from incommensurate CDW (ICCDW) to a nearly commensurate CDW (NCCDW) between 100 and 220 K, with relationship of temperature sweep direction in as-prepared few-layer 1T-TaS₂. The memristive I-V characteristics and hysteretic current-temperature curves can be caused by these phase transitions. It can be seen from Figure 2E

that 13 Ta atoms form a large satellite cluster in the CCDW phase, as zoomed-in in the inset. The adjacent NCCDW phase has a hexagonal arrangement originating from the CCDW domain, which is transformed into the ICCDW phase after further heating.

3.2 Artificial neurons

Neurons, consisting of cell bodies, axons, and dendrites, are the basic structural and functional units that transmit biological signals in the human body (Abbott and Nelson, 2000; Gerstner and Kistler, 2002) as listed in Figure 3A. A neuron receives signals from the anterior neuron *via* the dendrites and then transmits them to the posterior neuron *via* the axon. The cell body of a neuron determines its electrical response (i.e., the opening/closing of ion channels) according to the signal of excitatory or inhibitory potentials. Figure 3B shows the membrane potential in neurons with an excitatory or inhibitory potential. When the membrane potential is greater than the threshold, the ion channel opens and an action potential (spike) is generated. The generated action potential causes a continuous potential difference in the neuron so that the signal is transmitted to the axon and the neuron releases ions externally, returning to its initial state (resting state). When the membrane potential is less than the threshold potential, no action potential is generated and the signal charge escapes and returns to the initial state (Bear et al., 2016).

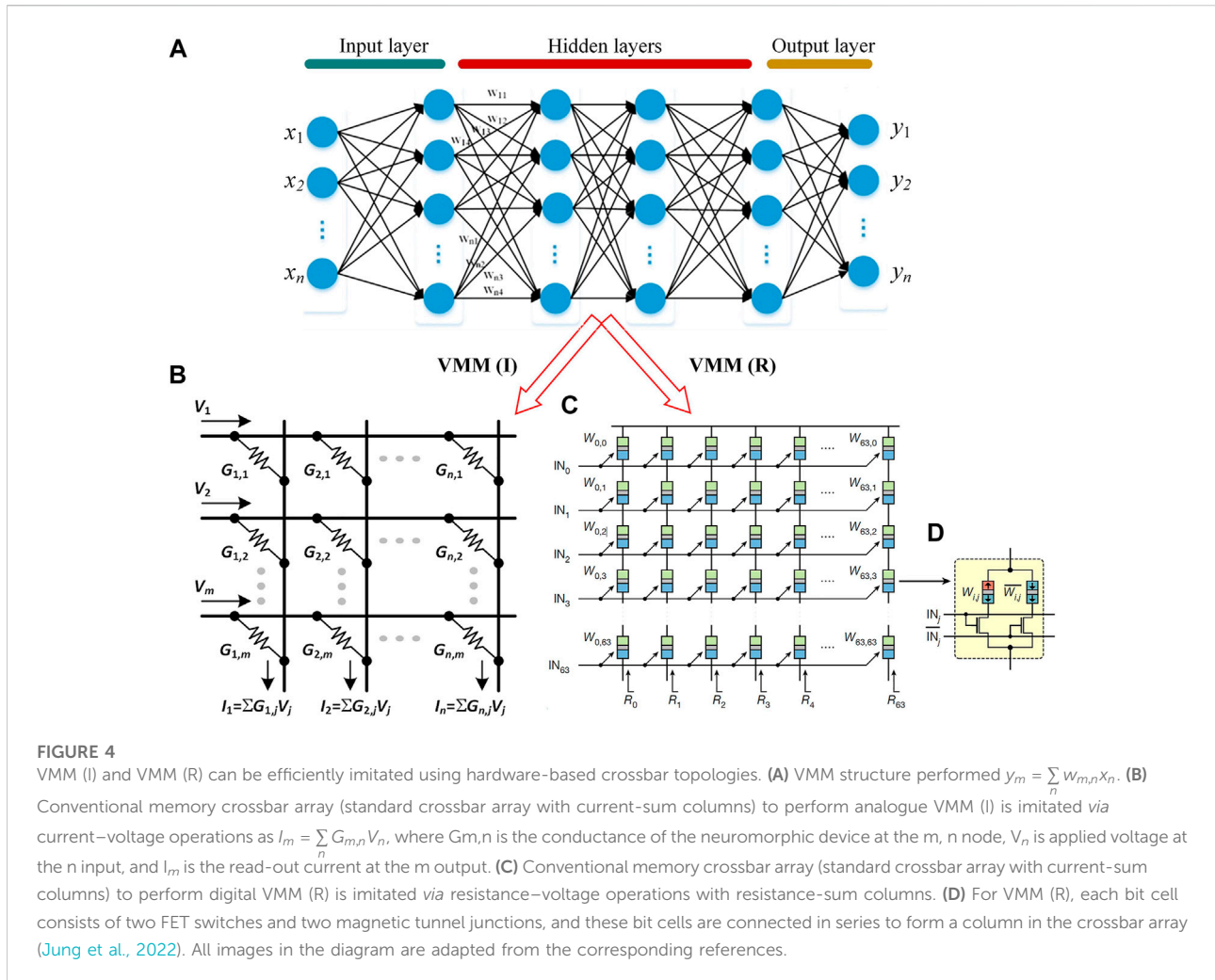
Mimicking this set of behaviors of biological neurons is a key for implementing artificial neurons (Zhang X. et al., 2017). Various models, such as Hodgkin–Huxley, Izhikevich, and integrate-and-fire models, have been proposed to explain the behavior of neurons and to implement artificial neurons (Sangwan et al., 2018; Wang et al., 2019a; Chen et al., 2019; Yin et al., 2019; Lee et al., 2020; Li et al., 2020; Wang H. et al., 2020). The leaky integration emission (LIF) model is the most widely used of these models (Figure 3C left), which assumes that the sub-threshold membrane potential dynamics of a neuronal membrane potential is similar to that of a resistor connected in parallel with a capacitor. The LIF model describes the behavior of the spiking nervous system very well (Yang et al., 2020). In particular, the LIF model, made from a leaky capacitor, adds the current from the synapse to the leaky current, helping to bring the neuron to a resting state. However, the excessive size of the capacitor increases the size of the neuron devices and its high power consumption severely limits its application.

The transistor circuits including comparators, summing amplifiers, and Schmitt triggers are employed in neuron models to achieve spiking behaviors efficiently. The right side of Figure 3C presents the configuration for the conventional summing and thresholding neuron (Chowdhury et al., 2017; Jiang et al., 2018), where the input current is accumulated and the corresponding voltage signal is sourced to the comparator by the

summing amplifier. If the output voltage of the amplifier is higher than the threshold (Du et al., 2021a), then the comparator will generate a voltage spike to the next layer of neurons. Despite the improvement over the previous one, it still suffers from size and energy consumption, so memristors have been suggested for the realization of artificial neurons (Tang et al., 2019).

The memristor design can be an alternative approach for artificial neurons (as listed in Table 2). The first amplifier with a memristor is not only to scale the output voltage, but also to implement the sigmoid activation function, in which the reconfigurable resistor of the memory is used to control the feedback gain, and the second amplifier is used to invert the output. Although memristors based on 2D materials have been used to implement neurons, the limitations in the preparation of volatile threshold switching devices mean that neuron devices have been less studied than synapse devices. The inherent characteristics (such as diffusive dynamics and the interfacial energy of the metal/vacancy species) (Valov et al., 2011) of component materials contribute to the performance of memristors, which operated with the filamentary mechanism. The characteristics of volatile resistive switching are usually required for most neurons based on memristive materials. Hao et al. (2020) verified that volatile resistive switching is possible in MoS₂-based neural components as listed in Table 2. In the case of CVD-grown MoS₂ in contact with Ag and W electrodes, the channel width of MoS₂ was 1.5 nm. Since both Ag and W electrodes are connected to the switching layer, the device only achieved stable volatile switching behavior by continuously adjusting the channel width of MoS₂ to 500 nm. In a vertical memory consisting of Ag/MoS₂/Au (Dev et al., 2019, 2020) as listed in Table 2, volatile/non-volatile properties were shown to be thickness-dependent. Thicker MoS₂ exhibits volatility threshold switching behavior, suggesting that thicker MoS₂ (~ 20 nm) can be used as a neural component. Here, this critical behavior of a medium-frequency artificial neuron firing as a function of the potential formed on it is implemented with the v-MoS₂/graphene threshold switching memristor (TSM) (Kalita et al., 2019) as listed in Table 2, and the conceptual scheme is listed on left of Figure 3D. The v-MoS₂/graphene neuron integrates the received input signals with the help of a capacitor. The capacitor consolidates the charge, and once the voltage across the capacitor increases above the TSM threshold, the neuron kicks in and produces an output spike. The schematic diagram of the v-MoS₂/graphene device is shown on the right of Figure 3D. It consists of CVD-grown monolayers of graphene, wet transferred onto a Si/SiO₂ substrate (Chan et al., 2012), and then patterned on graphene to grow v-MoS₂. Nickel contacts are deposited on graphene and v-MoS₂. An optical image of the device is shown on the right of Figure 3D.

Zhang et al. (2020) have designed a memristor-based artificial spiking somatosensory system (on the left of Figure 3E), which is a two-terminal sensor device and a compact oscillator, in which the special oscillator serves as the

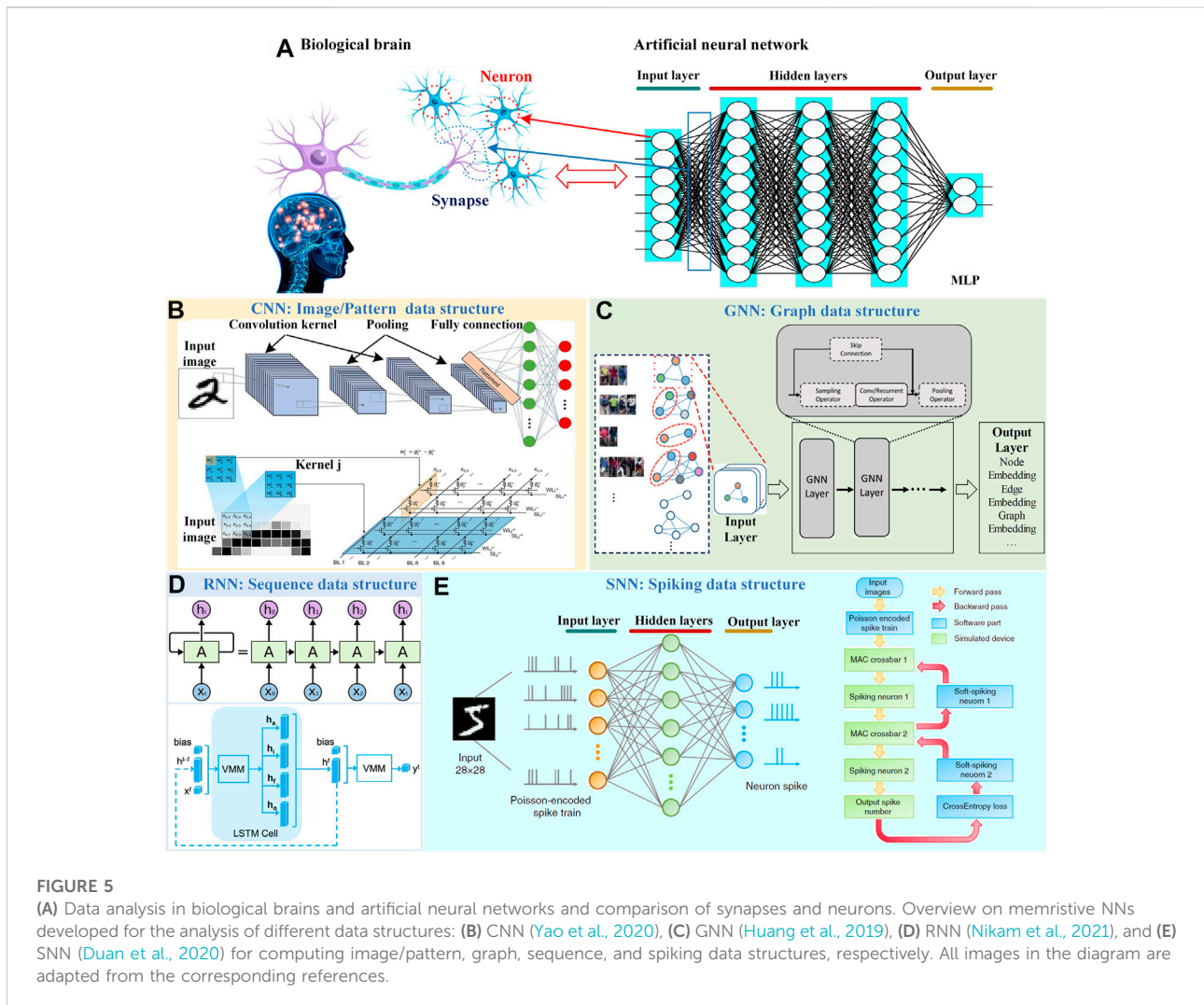


artificial spiking afferent nerve (ASAN) and contains two passive components: a resistor and a niobium oxide (NbO_x) memristor. In biological systems, the firing rate of afferent nerves increases with increasing input intensity whenever the intensity of the input stimulus exceeds the threshold of the afferent nerve (Sivaramakrishnan et al., 2004; Lin et al., 2006). However, when the stimulus intensity is very high, the firing rate decreases due to protective inhibition of neuronal cells to prevent neuronal death (Stetler et al., 2009). In this working artificial body sensing system, the analogue input voltage signal is generated by the sensor device and NbO_x ASAN can convert the voltage intensity into the corresponding spike frequency. The generated spikes will then be transmitted to NNs for further processing. The device has a titanium nitride (TiN) top electrode, a NbO_x switch layer, and a polysilicon bottom electrode. Here, the polysilicon bottom electrode with its low thermal conductivity is specifically designed to reduce the threshold current. Cross-sectional transmission electron micrograph (TEM) of the device structure is shown on the right of Figure 3E. A circular region of NbO₂ crystals with a diameter

of approximately 8 nm can be observed in the channel region at a close range (on the right of Figure 3E).

3.3 Vector–matrix multiplication

As a row-by-row vector generation process, vector–matrix multiplication (VMM) aims to generate new vectors using existing vectors, where each element of the new vector is generated by weighting and summing each row of the matrix using the elements of the vector as coefficients. In the BIC system, VMM is the basis for NN algorithms, which can be imitated by hardware-based NNs for efficient learning processes. For NNs, hidden layers are used to interconnect input layers (which can be pixel points or data values in the image) and output layers (as shown in Figure 4A), and each layer is interconnected with nodes to complete the information transfer. Therefore, the NN relying on VMM can be applied for inference or prediction as well as learning. The connectivity of nodes, that is, synaptic weights, can be modulated.



In terms of hardware, VMM based on crossbar arrays can be effectively implemented in synaptic devices using Ohm’s law and Kirchhoff’s law. However, since the conductivity of a synaptic device is always a positive value, the negative weight cannot be effectively implemented. In a typical approach, the conductance values of the two synaptic devices are subtracted to express a negative weight (Burr et al., 2015; Choi et al., 2019). Another method is weight shifting, where the median of the weight values is shifted from zero to a positive region so that all weight values become positive. Based on positive weight values, this can be expressed directly in terms of the conductivity of the synaptic apparatus (wij) (Han et al., 2021). The VMM performed with $y_m = \sum w_{m,n} x_n$ (Figure 4A) is imitated. Input signals (e.g., voltage vector) are delivered to each junction of different conductance, where a current as the multiplication result is generated according to Ohm’s law ($I = VG$, where G is conductance). The current at each column follows Kirchhoff’s current law, which allows one to obtain the summation of the

multiplications (currents of each junction) by measuring the total current out of each column ($I_j = \sum V_i G_{i,j}$) (Figure 4B). The abovementioned is the main mechanism of conventional crossbar arrays with current-sum columns. But in some applications, such as magnetoresistive random access memory crossbar arrays, due to small low resistance and high resistance, the conventional approach would consume considerable power (Xia and Yang, 2019).

In these cases, the resistance-sum columns have been introduced into the crossbar array, still with the aim of obtaining dot products (Figure 4C). The architecture starts with a new bit cell design (a bit cell is an element at a row-column intersection). Each bit cell connects two paths in parallel, each consisting of an MTJ and a field effect transistor (FET) switch in series (Figure 4D). The FET gates in the left path are driven by the binary input voltage I_N ($V_L = 0$ V, or $V_H = 1.8$ V), while the FET gates in the right path are driven by a voltage complementary to I_N . The MTJ–FET path on the left

stores a synaptic weight W (R_L or R_H ; each is the sum of the MTJ and FET switching resistances), while the MTJ–FET path on the right stores a weight complementary to W . Then, the left or right path can be selected by IN, generating the resistance (R_L or R_H) of the selected path as a bit cell output (Jung et al., 2022).

4 Architecture level: Neural networks determined by data structure

In the data-centric era, NNs are applied to solve different data structures for different application-specific scenarios. The network architectures in BIC and layer configurations are subsequently tailored, depending on the data structures. Inspired by the biology of the human brain, NNs have been constructed and applied to imitate the biological properties of the human brain (Figure 5A). As aforementioned, the artificial synapses correspond to the neuronal ties between the layers of the NN and serve to enable the transmission of signals between layers, while the artificial neurons correspond to the blue units in every layer of the NN (Figure 5A).

4.1 Neural networks

From architecture point of view, the input layer and the output layer are indispensable in the NN, between which are the hidden layers. One typical structure of the hidden layer is the full connection (FC) layer as shown in Figure 5A. The N neurons connected to the inputs are arranged vertically in a single line as the first layer (input layer), the same number of neurons are arranged vertically in a single line as the second layer (first hidden layer), and similarly there are M neurons arranged horizontally in a single line as the third layer (i.e., as the second hidden layer), and the last layer is the output layer. Each input neuron in the first layer and the corresponding output neuron in the second layer are connected to each other by synapses. It can be referred to as the SLP with only one hidden layer or the MLP with multiple hidden layers. The DNN is used for specifying the machine learning (ML) in general based on the previously described NN architectures with multiple hidden layers. In addition to the FC layers, different layer constructions can serve as hidden layers for diverse application scenarios. In this subsection, we discuss the following four representative NN architectures for BIC systems: CNN, GNN, RNN, and SNN.

4.1.1 Convolutional neural network

A feed-forward NN consisting of one or more convolutional layers and pooling layers is termed as the CNN, which responds to surrounding units in the coverage area. The CNN is excellent for processing image/pattern data structures. In addition, the input layer, activation function layer, associated weights, and FC layer are indispensable for building a complete CNN (Figure 5B).

For the hardware implementation of CNN architecture, the input layer consists of the same number of neurons as the pixels of the input image/pattern. N different convolution kernel groups form the convolution layer, and the size and number of which also depend on the input image/pattern. The convolution kernel can be shared during data processing, which is useful for the small size of NN, and the input image can still retain the original positional relationship through the convolution operation. All the neurons are formed by synaptic connections between them. First, the images in the data set are sorted and numbered. The convolution operation is then performed by sliding continuously with a fixed step size and computing the sum of weights between the shared local kernel and the input blocks generated by the input layer. The output is the first convolution layer, which is then sub-sampled by the first pooling layer, at which point the first round of convolution operations is complete. The subsampling result from the subsequent pooling operation becomes the input data for the second convolution operation, and then loops the structure above. The number of convolution layers depends on the size of the input data. For memristors crossbar computing architecture, the VMM is the basic algorithmic operation in the convolution procedure. The convolution needs mapped into VMM by converting high-dimensional convolution into low-dimensional VMM. Then, the low-dimensional convolution can be implemented using matrix multiplication by converting kernels (Zhang and Hu, 2018). Then, a kernel is mapped throughout to the corresponding positive and negative weight rows, the later pooling layers are expanded and then these are expanded into vectors. Thus, the input of highly dimensional and complex data is effectively implemented as dimensional compression, that is, dimensionality reduction. Then, the data with reduced dimensionality can be transferred to the final FC layers. Then, the value of the weight summation is fed as the input of the softmax function to calculate classification probability. This is the complete execution process of a CNN. The convolution operation based on memristor crossbar has two implementation methods. One is that the input feature maps in turn are input into sliding windows based on compact memristor crossbar to obtain the output feature map (Yakopcic et al., 2016). The other is that an entire feature map is input to a sparse crossbar (Yakopcic et al., 2017), but lots of redundant memristors may exist and it is challenging to make the conductance of the same convolution kernel uniform.

4.1.2 Graph neural network

The GNN is an NN that acts directly on the graph, which consists of an input layer, a hidden layer, and an output layer (Figure 5C). The GNN is capable of processing graph data structures. First, given a graph, the nodes were initially converted into recurrent units and the edges, into feed-forward NNs. For example, we defined the features for any social network graph, where the features could be age, gender,

address, dress, etc. The nodes connected by each edge may have similar characteristics. This reflects some kind of correlation or relationship between these nodes. Then, the nodes are labeled, all nodes are converted into recurrent units, and all edges contain simple feed-forward NNs. At this point, the node and edge transformations are complete, and the graph can pass messages between the nodes. Then, n times nearest neighbor aggregation (i.e., message passing) is performed for all nodes. This is because it involves pushing messages (i.e., embedding) from surrounding nodes *via* directed edges around a given node. For a single reference node, the nearest neighbor nodes pass their messages (embedding) to the recurrent units on the reference node through the edge NN. The new embedding update with reference to the cyclic unit is based on using the cyclic function on the sum of the output of the edge NN, with cyclic embedding and nearest neighbor node embedding. Then, the embedding vectors of all nodes are summed up to obtain the graph characterization parametrization. Last, one can skip H altogether and go directly to higher levels or use the parametrization to characterize the unique properties of the graph as well. Most GNNs will choose “convolutional” layers, as those in CNNs (Yujia et al., 2016; Hamilton et al., 2017). The “convolution” operation in GNN can be roughly divided into two phases (Yan et al., 2020). The aggregation phase aggregates nodes’ information from their multi-hop neighbors by pointer-chasing operations. This phase incurs intensive random memory accesses. The handling phase feeds the aggregated features into an NN to generate new features. Both computation and aggregation are regular in this phase. This process is executed in parallel on all nodes in the network, since the embedding of the $L+1$ layer depends on the embedding of the L layer. Therefore, there is no need to “move” from one node to another to pass messages. It happens that the GNN directly propagates the graph structure rather than treating it as a feature and maintains a state that can represent information from a human-specified depth.

4.1.3 Recurrent neural network

NNs that can process sequential data using recurrent processing units are called RNNs, which consist of an input layer, a hidden layer (containing recurrent kernels), and an output layer. The RNN is capable of processing time sequences and image/pattern data structures. The hidden layer of a conventional RNN is a tangent function that acts as a memory unit for the activation function, while the output W at moment $t-1$ is re-weighted as input and re-entered into the neuronal activation function to jointly act on the output result at the current moment t . As the time step increases, the values within the early memory gradually increase to astronomical numbers, which can lead to the more long-term memory having a greater impact on subsequent output, while short-term memory is rather less influential, and such a structure can cause gradient disappearance and explosion problems. Later,

a special type of RNNs emerged, namely, the long–short term memory network (LSTM). The LSTM consists mainly of an oblivion gate, an input gate, an output gate, and a memory cell. Its hidden layer is a self-connected linear cell, also known as a constant error carousel (CEC). CEC protects RNNs from vanishing and exploding gradient problems of the traditional RNN.

For the hardware implementation, all the neurons between the layers in the hidden layer are interconnected, but the neurons between the same layers are not. Each neuron in the hidden layer of an RNN introduces a recurrent synapse that stores the neuron’s output at moment t as its own input at moment $t+1$, and similarly, the output at moment $t+1$ will serve as input at moment $t+2$, and so on. The RNN achieve information transfer through it (as shown in Figure 5D). The structure of RNN is an iterative process and the weights are shared, that is, the input x at different moments uses the same weight matrix for each part each time, which can reduce the number of parameters and thus the complexity of the computation. The key point is the recurrent kernel, which remembers the features of the last hidden layer output and passes it to the next input. Each of the inputs and outputs can be of indefinite and unequal length. RNNs are very effective for data with sequential properties, which can mine the temporal information as well as semantic information in the data. This ability of RNNs to process sequential data recursively is exploited to make a breakthrough in deep learning models for solving problems in natural language processing (NLP) fields such as speech recognition, language modeling, machine translation, and temporal analysis.

4.1.4 Spiking neural network

The SNN is one of the third-generation NNs, which is to process the data in a more biological fashion in comparison to other aforementioned NNs. The structural units for transforming data into spiking sequences are included in addition to the input, hidden, and output layers (Figure 5E). The SNN is capable of processing spike sequence data structures.

SNNs process information using the timing of signals (pulses). In contrast to actual physiological mechanisms, SNNs almost universally use an idealized pulse generation mechanism. First, a number of pulse spike sequences are fed to the neuron, and then these spike trains are fed into the memristor crossbar and converted into weighted current sums through the columns. A row of transimpedance amplifiers can be used to amplify and convert the currents to analog voltages of (-2 and 2 V). The neurons can then integrate the analog voltages and generate spikes when reaching the firing threshold, which propagate to the next layer for similar processes. At last, the spiking numbers of output neurons are counted, and the index of most frequently spiking neuron is taken as the prediction result. In the brain, communication between neurons is accomplished by propagating sequences of action potentials (also called pulse sequences transmitted to downstream neurons). What it

TABLE 3 Summary of the features in data structures which determine the application of NNs.

Data type		Data feature	Time-relatedness	Neural network	Optimization	Application	Data analysis	
Unstructured data	Discrete data structure	Discrete (discrete domain)	Same data set with different sizes	No	MLP	Several local optimization	For example, robotic controls; drug discovery; language translation; and signal integrity problem	Parallel (synapse depends on data types)
	Non-Euclidean data structure	Graph (graph domain)	Differed data set with different sizes	No	GNN (graphic NN)	Global optimization	For example, auto random image recognition; molecular drug discovery; molecular fingerprint analysis; and phase transitions in glasses	Parallel (synapse depends on nodes)
Structured data	Euclidean data structure	Image/pattern (image domain)	Same data set with same size	No	MLP, CNN, and RNN	Local (MLP, CNN), global (RNN, CNN)/	For example, recognition cat and dog	Parallel (synapse depends on pixel)
	Sequential data structure	Sequential (time domain)	Same data set with different sizes	Yes	RNN (recurrent NN), training among somas	Global optimization	For example, stock market, relationships among stocks A/B/C matters	Serial
		Binary pulse sequence (spatiotemporal domain)	Same data set with different sizes	Yes	SNN, event sequence is important	Local/global optimization	For example, associated learning process	Serial (synapse depends on data types)

enhances is the ability to process data in spatiotemporal, spatial meaning that neurons only connect with nearby neurons so that they can process input blocks separately. Temporal means that pulse training occurs over time so that information lost in the binary encoding can be regained in the temporal information of the pulse, which allows to process temporal data naturally; and SNN can handle time series data without additional complexity. The neuronal units in SNNs are active only when spikes are received or sent, thus making them energy efficient. It is therefore event-driven, thus allowing it to save energy.

4.2 Data structures

In this section, we discuss five representative data structures in application-oriented scenarios applied to BIC systems: image/pattern data structure, graph data structure, sequence data structure, spiking sequence data structure, and discrete data structure.

4.2.1 Image/pattern data structure

Image/pattern data are a collection of grayscale values for each pixel expressed as a numerical value and are a type of structured data, also known as Euclidean data, and the data structure is high-dimensional, complex. In the practical application of the BIC system, there are many image/pattern data structures being expected to be processed by suitable NNs.

In addition, there are also many other data types that cannot be processed into image/pattern data types, such as photonic devices that have freeform geometries, and these data cannot be parameterized by a few discrete variables. But it is possible to convert these data into 2D/3D images/patterns. It will be traded off with other performance metrics to realize all the features of the input (full connection), so each neuron is connected to only one local region of the input data, called the receptive field, and the size of the connection is equal to the depth of the input quantity in the depth direction. Image data types are effectively processed using a set of convolutional layers in series that can extract and process spatial features (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015; Szegedy et al., 2017). The CNN can extract a large number of local close features and combine them into high-order features. For this high-dimensional and complex image data, the downscaling of image data is achieved through 2–3 layers of convolutional operations by kernel sharing and convolutional operation in the CNN, and finally the classification probability is calculated for low-dimensional data operations. Therefore, image/pattern data structure is regarded to be more suitable for CNN architecture to process (as listed in Table 3).

The CNN with five layers, which contains two convolutional layers, two pooling layers, and one FC layer was applied for recognition of the MNIST (LeCun et al., 1998) digit images (Yao et al., 2020). The max-pooling and rectified linear unit (ReLU) activation functions are employed and obtained the accuracy of

96.19%. The results show that more than two orders of magnitude has better power efficiency and one order of magnitude has better performance density compared with Tesla V100 GPU. The CNN architecture for the model was employed to recognize the image data (LeCun et al., 2015; He et al., 2016). In addition, image-processing-related tasks, such as image segmentation and object detection (Ren et al., 2015), image classification (Simonyan and Zisserman, 2014), video tracking (Fan et al., 2010), and NLP (Karpathy and Fei-Fei, 2015) are also still inseparable from CNN architecture.

4.2.2 Graph data structure

In computer science, a graph is a data structure consisting of two parts: vertices and edges. A graph can be described by the set of vertices and the edges it contains, which is a data structure that models the relationship between nodes and nodes. Graph data consist of nodes and undirected edges with label information, which are also the only non-Euclidean data in ML. Graph is the disorder, and can express starting and ending points without clear information. Graphs can represent many things—social networks, molecules, etc. Nodes can represent users/products/atoms, and edges represent connections between them, such as following/usually buying/keying at the same time as the connected product. A social network graph may look like this, where the nodes are users and the edges are connections. There are two main types of graphs: directed graphs and undirected graphs. In a directed graph, there is a direction of connections between nodes; in an undirected graph, the order of connections does not matter. A directed graph can be either unidirectional or bidirectional (as listed in Table 3).

Therefore, graph data structures are suitably processed in GNNs (Henaff et al., 2015; Niepert et al., 2016; Veličković et al., 2017), which analyze and operate on aggregated information between neighboring nodes in each layer. The node classification is a typical application with GNN architecture, each node in the graph is associated with a label and we are required to predict the label of unlabeled nodes. GNNs have been applied to broad aspects, including the modeling of molecular drug discovery (Torng and Altman, 2019), molecular fingerprint analysis (Duvenaud et al., 2015), and phase transitions in glasses (Bapst et al., 2020). In addition, graph network architectures are highly specialized depending on the actual application, such as graph generative networks (Ma et al., 2018), graph recurrent networks (Li et al., 2015), and graph attention networks (Veličković et al., 2017).

4.2.3 Sequence data structure

In addition to discrete data and images/patterns data, there are also large-volume sequential data structure and sequential problems. Sequence data are the data collected at different points in time and reflect the state or extent of change of a thing, phenomenon, etc. over time, such as time series data structures

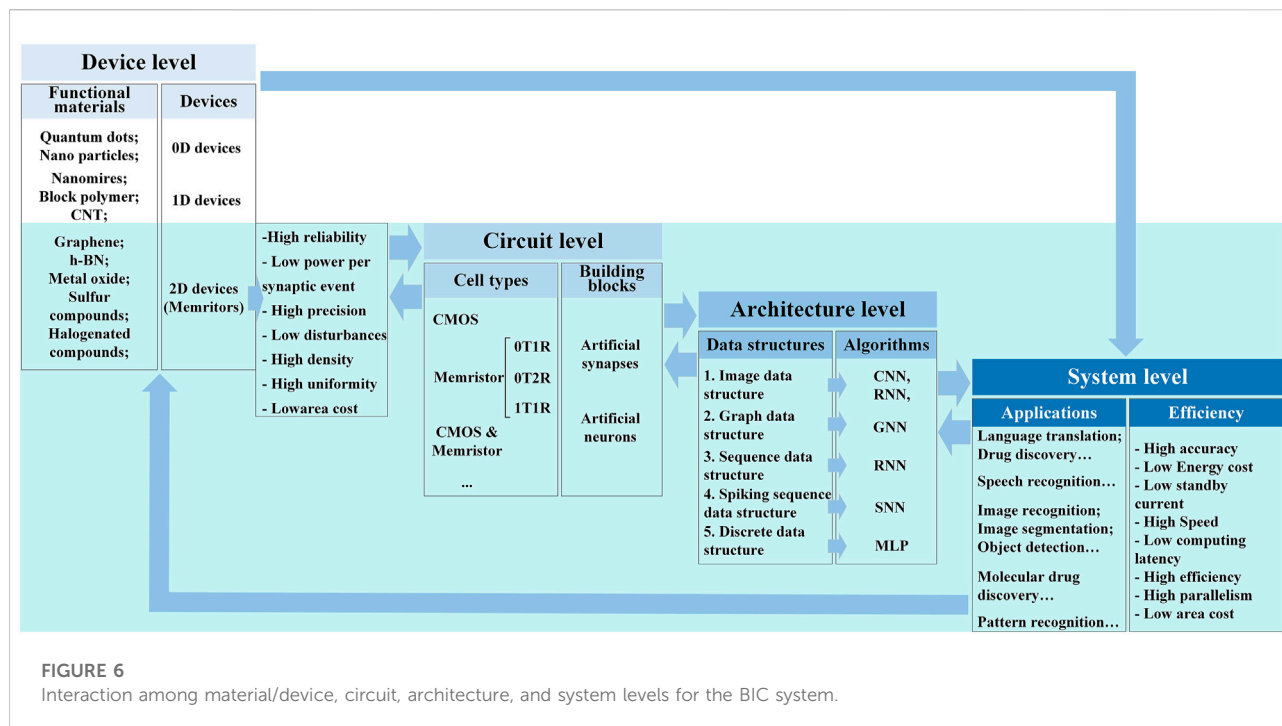
and text sequences, but there is a temporal correlation between data and a dependency between, before, and after for series data. In addition, there are variables and responses among mathematics and physics, for example, in dynamic electromagnetic systems; when the discrete time steps are small enough, these continuous electromagnetic change phenomena can be represented by discrete time series without going general. Moreover, both the input and output signals are time series, and the output at a given time depends not only on the input at the current moment, but also on the state of the device at the previous time step (i.e., the electromagnetic variation of the device).

For example, the rise and fall of a stock requires prediction of the data at the next moment, so the output depends not only on the input, but also on the memory (that is, the output at the previous moment can act on itself again). RNNs feed the network outputs back into the input layer, maintaining a memory that accounts for the past state of the system, which makes them ideally suited to model time-sequential systems (Graves, 2013; Weston et al., 2014; Luong et al., 2015) as listed in Table 3.

4.2.4 Spiking sequence data structure

Spiking sequence data structures are discrete spike signals containing both spatial data information between layers and temporal data information, that is, spatiotemporal data. In the human brain, when any of the body's senses are stimulated, the neurons in the human brain produce an impulse signal containing action information. There are two forms of impulse signals, one is directly spiking sequence data, and the other is the signals collected from the environment, which are usually in the continuous and analog domain, and need to be transformed into spiking sequence data first to serve as the inputs to SNNs. These individual pulses are temporally sparse, each with a high information content and approximately uniform amplitude. Thereafter, the spiking sequence data structures are well-suited to be processed by SNN architecture as listed in Table 3.

In 1997, the SNN has been proposed as the third-generation NN (Maass, 1997), and the SNN has been widely studied. A million spiking neuron into circuit with a scalable communication network and interface were integrated to achieve multi-object detection and classification tasks (Merolla et al., 2014). The Loihi chip was applied to solve LASSO optimization problems with over three orders of magnitude superior energy-delay product compared to conventional solvers running on a CPU iso-process/voltage/area (Davies et al., 2018). This provides an unambiguous example of spike-based computation, outperforming all known conventional solutions. SNNs have become the focus of many recent applications in many areas of pattern recognition, such as vision processing, speech recognition, and medical diagnostics.



4.2.5 Discrete data structure

Discrete data are a class of countable data that are numerically independent and specific in nature for certain specific values and lack correlation between the data. Typical discrete parameters are, for instance, the device geometry such as the height, the width and the period, or the permittivity and permeability of a material. In addition, there are also many properties of devices/objects described as discrete parameters, which include device efficiency, quality factor, band gap, and spectral response sampled at discrete points. Owing to lacking correlation, they cannot be effectively studied by partial or related parameter analysis and feature extraction. Discrete data structures naturally interface with MLP (as listed in Table 3).

DNN was applied for the signal integrity problem. The characters in signals are classified by a large-scale memristor-based DNN in Shin et al. (2020), and a three-layer NN was constructed and presented the classification accuracy of 99.4% by using ReLU as the activation function. Using DNN training and real measurements, the magnitude of the impact of interconnection parasitic of memristor arrays problems (e.g., IR drop, crosstalk or ringing) on signal integrity was determined. Similar research works have been carried out (Chen et al., 2015; Lee and Kim, 2019). In addition, research works on robotic controls (Abbeel et al., 2010) and drug discovery (Lavecchia, 2015) to image classification (Krizhevsky et al., 2012; Liu et al., 2016; Adam et al., 2018; Tang et al., 2020; Liu and Zeng, 2022) and language translation (Wu et al., 2016) have also been performed. These algorithms will only get more

powerful, particularly given the recent explosive growth of the field of data science. The memristor-based DNNs can be applied for XOR operation and digit recognition on the MNIST (Modified National Institute of Standards and Technology) (LeCun et al., 1998) data set. The classification accuracy is up to 96.42% (Liu and Zeng, 2022).

5 Discussion (system level aspect)

The development and implementation of the BIC system is an interdisciplinary work, which is determined by the design in the device level, circuit level, up to architecture level and system level as demonstrated in Figure 6. Upon the systematic review on the aforementioned design levels, in this section, the interactions between the system level and device/circuit level and between the system level and architecture level are discussed.

5.1 Cross-comparison on the device/circuit level

For the BIC system design, the choice of materials and devices in the integrated circuit significantly influences the construction in the architecture and system levels, especially in terms of power consumption, area density, computing speed, and system accuracy.

In the device level, the ideal functional material for integrated devices in BIC possesses the features with easy fabrication possibility, excellent optical/electrical properties, high stability, microstructural tunability, and controlled charge and promising innovation. As reviewed in [Section 2](#), the 0D materials have excellent optical properties and can be fully valued in photonic systems, but their stability is poor and the selection is narrow, so the variety of 0D devices are also limited. Inevitably, 0D devices face the context of fabrication challenges, especially among 0D QDs, where there are obvious difficulties in obtaining quantum coherence. Furthermore, in high-mobility semiconductors, the number of quantum states is critical ([Altaisky et al., 2016](#)). An important point of 1D materials is the general bottom-up growth characteristics, with metal or semiconductor behavior, which depends on its chiral vector, but single 1D materials have always been a challenge in application and alignment. So, it is limited to application in promising 1D materials with the lack of NVM characteristics and preventing the inherent transverse geometry of the dense cross switch array. The 2D materials present the following unique features: first, the layer-by-layer exfoliation of 2D bulk materials can be achieved due to weak van der Waals forces between adjacent layers; second, the charges can be precisely controlled by the gate voltage to eliminate influences; and third, the promising innovation can be achieved by controlling the energy band structure. For instance, the 2D memristor devices with excellent properties as shown in [Figure 6](#) can be served as ideal device candidates in the circuit level in the BIC system. Furthermore, the integration of low-dimensional materials (0D–1D, 0D–2D, and 1D–2D) ([Sangwan and Hersam, 2020](#)) is regarded as an efficient method to avoid its own problems.

In the circuit level, 2D memristor devices are considered as an example; the ideal memristive devices for mimicking the functional behavior of synapses and neurons in the NN should possess low switching energy, small feature size, high device conversion speed, low device variability (high reproducibility), high endurance, and symmetrically programmable conductance states (for possible linear weight training), especially with the ability of VMM as demonstrated in [Figure 6](#). Particularly, the device reliability demonstrates a strong impact on accuracy in BIC. For example, the artificial neurons and artificial synapses are needed for processing and transmitting the signals in the BIC system. Taking 2D memristors as an example, as reviewed and illustrated in [Table 2](#) in [Section 2](#), the 2D memristors have many options for implementing synapse devices, but very few for neuron devices, which is closely linked to the choice of functional materials used as the switching layers. It proves that the working principle of devices cannot be separated from the structure and properties of the material. The application of the same materials to devices with different working principles results in very different sets of voltages and currents, which also determine the magnitude of power

consumption. Therefore, the role of functional materials is indispensable for the design of different types and switching mechanisms of the memristive devices. It is the excellent characteristics of the devices themselves that enable the possible functional properties of the artificial neuronal circuits. Note that, despite in general, the longer retention time is favored in memory devices, the requirements of retention time can be strongly dependent on the targeted building blocks in the BIC system. The artificial synaptic devices based on 2D MoS₂/graphene memristors were constructed and exhibited essential synaptic behaviors, especially excellent retention characteristics of 10⁴ s ([Krishnaprasad et al., 2019](#)), which is attributed to the fact that synapses present long data retention after electrical stimulation. However, as for the artificial neuron based on a diffusive memristor, after generating the signal, it quickly resumes its initial state to wait to respond to the next signal. Thus, a shorter retention performance should be presented ([Wang et al., 2018](#); [Du et al., 2021b](#)).

The energy cost is crucial in the performance evaluation of the BIC system, especially the energy consumption generated per synaptic events. In numerous reports, the energy consumption of CMOS-based artificial synapses is at ~nJ per event level ([Painkras et al., 2013](#)). However, it is easy for memristive artificial synapses to reach several ~pJ per synaptic event ([Yu et al., 2011](#); [Jackson et al., 2013](#); [Li et al., 2022](#)), or even several hundreds of ~fJ ([Xiong et al., 2011](#); [Pickett and Williams, 2012](#)), which is close to those of the biological brain. As reviewed in [Section 2](#), the application of different materials to the devices shows different energy consumptions, which is due to the properties of the materials themselves. Particularly with 2D materials, which are used in memristor devices, very low energy consumptions can be obtained consistently [h-BN ([Vu et al., 2016](#); [Shi et al., 2018](#)), MoS₂ ([Knoll et al., 2013](#)), HfSe₂ ([Li et al., 2022](#))]. Therefore, for memristor devices, the choice of switching layer material has an inestimable impact on the energy consumption of the device and is also a potential direction. In addition to the devices category, considering the huge number of synapses in the Mem-BIC system, such as 10⁵ synapses in the application ([Kornijcuk et al., 2016](#)), the energy consumption of memristive synaptic operations has been significantly reduced by several orders of magnitude compared to conventional CMOS technology. Furthermore, 2D memristor devices optimize the performance of parallel VMM computing in terms of power consumption (include operating current/voltage), and number of states, providing a superior platform for the future BIC system.

5.2 Cross-comparison on the architecture level

In addition to the device and circuit levels, the architecture level plays a crucial role in the BIC system as the NN architecture is a direct reflection of the processing requirements in big data

TABLE 4 Comparison of performances (accuracy and power consumption) of different NNs (different sizes) depending on different data structures.

Data structure	NN	Accuracy (%)	Power (mW)	Size (no. layer)	Application	Reference
Image	CNN	96.2	7.44	7	MNIST image recognition	Yao et al. (2020)
Pattern	CNN	94.6	306.70	4	Image classification	Cai et al. (2019)
Pattern	MLP (DNN)	96.5	2.180	4	Pattern recognition	Liu and Zeng (2022)
Image	MLP (DNN)	94.6	100.60	5	Pattern recognition	Liu et al. (2016)
Discrete sequences	MLP (DNN)	99.4	N/A	5	Signal integrity	Shin et al. (2020)
Pattern	RNN (LSTM)	96.1	312.40	3	Pattern recognition	Smagulova et al. (2018)
Pattern	RNN (LSTM)	99.0	255.80	3	Pattern recognition	Adam et al. (2018)
Time sequences	RNN (LSTM)	95.0	105.90	3	Forecasting problem based on the Keras model	Smagulova et al. (2019)
Time sequences	RNN (RC)	99.6	0.05	3	Temporal signal processing	Zhong et al. (2021)
Time sequences	RNN (RC)	99.2	0.30	3	Temporal data classification and forecasting	Moon et al. (2019)
Spike sequences	SNN	89.1	N/A	4	MNIST classification	Kim et al. (2021)
Spike sequences	SNN	90.0	2.90	3	Handwritten digit recognition	Nowshin et al. (2022)

flow. Despite the optimized energy cost per switching event or synaptic event in the device/circuit level, more significant power gain can be found in the architecture level in NN architectures with comparable data size/representation. Note that the overhead design for controlling memory blocks, that is, the circuit design for data movement will further influence the energy dissipation in the architecture level.

Table 3 summarizes the features of data structures for mostly applied data sets in different application scenarios. For unstructured data, limited NN structures can be used to process it directly, showing inconsistencies in the data set and data size. Specifically, discrete data structures lack tight correlation between data and have no timeline. In the field of neuromorphic computing, the graph is the only non-Euclidean data structure, which presents different data sets and data sizes. For structured data, there are strong correlations between data, such as image/pattern with positional correlations, and sequential data and pulse sequences with temporal and logical correlations. Therefore, it is possible to capture the main features of the data while processing this type of data, thus reducing the dimensionality of the data and simplified NN to obtain an efficient system. In particular, the types of data associated with time are becoming more and more abundant (such as brain signals, stock data, gross domestic product data, and business cycles), and the requirements for NNs will also increase.

Based on the aforementioned summary of different data structures and characteristics, MLP is suitable for discrete data lacking correlation between data and having no timeline in unstructured data. All outputs from the last layer are inputs to the next layer, and all data information has a logical relationship with adjacent layers. GNNs may belong to a class of NN architectures, especially tailored for this purpose, and thus are widely employed to process the graph, which is the only non-

Euclidean data structure in the field of neuromorphic computing. CNN, RNN, and SNN can be applied for processing structured data, including image/pattern with positional correlations, sequential data, and pulse sequences, with temporal and logical correlations. CNNs include unique internally sharable volume and pooling layers, which reduce the dimensionality of the data and simplify the NN. For RNNs, the unique feature is the temporal correlation, where the output of the previous moment can be fed back and applied to the current moment, which works well for data with strong temporal correlation (such as brain signals, stock data, gross domestic product data, and business cycles). The most obvious feature of the SNN is that it most closely resembles human brain signaling and processing and is ideally suited to impulse signals similar to those of the human brain. It is clear from Table 4 that structured data such as images/patterns are common in practice. There is also more than one type of NN used to process these data, including CNNs, RNNs, and their variants, while for discrete data and graph data, there are very few practical applications of BIC based on memristors.

After specifying NN models based on different data types, the desired behaviors for artificial synapses and neurons deserve more attention. The ideal memristive devices for artificial synapses and neurons should possess low switching energy, small feature size, high device conversion speed, low device variability (high reproducibility), high endurance, and symmetrically programmable conductance states. For distinct NN architecture models, which in general also correspond to different data structures, there are significant differences in architectures. Not only does the data input require more pre-processing units, but the structure is also adjusted to the complexity of the data during processing, which inevitably causes latency in the processing of the data. As NN requirements increase and structures become more complex, it

is inevitable that devices and circuits are added to the processing unit module, which directly challenges their feature size, device conversion speed, and device variability. For the same architecture, the data type is fixed, but the volume of data varies, so the number of NN layers is not the same. Latency will occur as data are processed in layers and passed between layers. The trade-off with other performances for devices and circuits naturally occurs. Consequently, NN models with distinct complexity and the volume of data should be considered with the desired behaviors for artificial synapses and neurons.

Furthermore, the BIC system is designed for dealing with the data-centric computing tasks. The ideal BIC system should demonstrate one or more excellent performance metrics such as high accuracy, low power consumption, low latency, and miniaturization to complete computing tasks. This requires a high level of system integration, minimal area, and high computational speed, which also implies good parallelism and depends directly on the design and requirements of the device and circuit levels. With short computation latency, the data flow should be transmitted in a very timely manner during system operation. Accuracy cannot be ignored in the practical application of BIC systems, as any other parameter that is well-maintained will serve the requirements of the final application, which places high demands on the reliability, retention, endurance, and uniformity of the circuit level and device level.

Table 4 summarizes the performances, including accuracy and power consumption, of the various applications using different sizes of NN for different data types. The accuracy based on CNN architecture is significantly the highest when dealing with the same data structure among them, which is due to the convolution and pooling operations in the structure, which drastically reduce the complexity of the data, thus guaranteeing high accuracy and low power consumption. However, the MLP itself is weak in terms of accuracy due to the structure of fully connected layers, which results in a large and complex amount of data, and its size (number of layers) is generally large. Little has been reported in terms of power consumption, which may be attributed to its structure and the amount of data that inevitably leads to excessive power consumption. It is worth noting that RNNs have relatively low accuracy when dealing with the pattern data structure, and the size of NNs is small, but the energy consumption is very large compared to that of CNNs and MLP. However, it is clear that when RNNs process time sequence data structures, the accuracy is high, the power consumption is small, and the size is also small. As the third-generation NN, the SNN has apparently the lowest accuracy in processing the spike sequence data structure, and little has been reported about power consumption; a recent study on it reported that energy consumption is very low and the size is the same as that of the RNN. Therefore, it has to be said that for the choice of NN in the architecture level, the input data structure plays a key role in the BIC system. In other words, the architecture level can improve

the efficiency of the system by determining the NN structure of the system through the type of data structure, but there is limited scope for optimizing the efficiency of the system in this way.

The area is an important efficiency metric of the system and depends on both the device and circuit levels, as well as on the NN architecture structure and the number of layers in the NN. The different devices such as transistors, memristors, or memtransistors directly determine the size of the area in design and production. Numerous studies have shown that due to their structure and characteristics, memristors can achieve better properties in smaller area sizes than transistors. Therefore, the same effect naturally exists for the circuit level consisting of devices. In addition, the area size of the circuit also depends on the functionality that the architecture is trying to achieve and the number of modules increases or decreases to a greater or lesser extent. As the architecture depends on the data structure, complex data structures naturally require more complex architectures and more layers of NNs, and the area is naturally larger. On the other hand, for smaller and simpler data structures, no complex architecture and NN numbers are required, and the area size is naturally reduced.

5.3 Challenges and perspectives

First, taking the advantages of layered controllable structures, charge, and energy level controllability, 2D materials will offer many opportunities and possibilities in current neuromorphic computing and next-generation neural network computing. However, there are challenges with high-quality two-dimensional films and their further stable modification by doping. So, 2D material preparation methods and more efficient doping are the promising direction. Aiming for high-performance 2D devices, small sizing and excessive drive currents due to high interfacial resistance are still challenging issues at present, as well as difficulties in the preparation of heterojunction arrays. There are therefore many opportunities for the control of multiple aspects of device structure contact surfaces and packaging, phase engineering, or self-alignment techniques and heterogeneous 3D and 2D structures. In the circuit level, these challenges and opportunities are also presented, where minimal size and low energy consumption due to low operating currents and voltages have always been the focus of our efforts. In the architecture level, although a variety of NN architectures exist for different data types, there are still significant challenges and opportunities for efficient execution of operations. At the system level, good energy efficiency and area are always important, but there are still many challenges based on the different levels mentioned earlier, which result in high energy consumption, low area efficiency, and low throughput. Consequently, these

challenges exist so that more opportunities are given to us to commit to them for the sake of optimizing with energy efficiency, minimal size, area, etc.

6 Conclusion

Consequently, the advances for the development of BIC systems with respect to the perspective of a system designer are reviewed in this work. For the four levels from the device technology level and circuit level, up to architecture and system levels, not only are the interactions among them discussed but the challenges and perspectives of each level are also elaborated in the end. Particularly, the NN architecture determined by the data structures centered on big data flows in application scenarios are sorted out meaningfully. This review can serve in the future for further development of the optimization of the BIC system design.

Author contributions

WW wrote the manuscript with support from ND, SK, and HS. All authors provided critical feedback and helped shape manuscript.

References

- Abbeel, P., Coates, A., and Ng, A. Y. (2010). Autonomous helicopter aerobatics through apprenticeship learning. *Int. J. Robotics Res.* 29, 1608–1639. doi:10.1177/0278364910371999
- Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nat. Neurosci.* 3, 1178–1183. doi:10.1038/81453
- Abdelouahab, M.-S., Lozi, R., and Chua, L. (2014). Memfractance: A mathematical paradigm for circuit elements with memory. *Int. J. Bifurc. Chaos* 24, 1430023. doi:10.1142/s0218127414300237
- Adam, K., Smagulova, K., Krestinskaya, O., and James, A. P. (2018). “Wafer quality inspection using memristive lstm, ann, dnn and htm,” in 2018 IEEE Electrical Design of Advanced Packaging and Systems Symposium (EDAPS) (Chandigarh, India: IEEE), 1–3.
- Alibart, F., Pleutin, S., Guérin, D., Novembre, C., Lenfant, S., Lmimouni, K., et al. (2010). An organic nanoparticle transistor behaving as a biological spiking synapse. *Adv. Funct. Mat.* 20, 330–337. doi:10.1002/adfm.200901335
- Alibart, F., Zamanidoost, E., and Strukov, D. B. (2013). Pattern classification by memristive crossbar circuits using *ex situ* and *in situ* training. *Nat. Commun.* 4, 2072–2077. doi:10.1038/ncomms3072
- Altaisky, M. V., Zolnikova, N. N., Kaputkina, N. E., Krylov, V. A., Lozovik, Y. E., and Dattani, N. S. (2016). Towards a feasible implementation of quantum neural networks using quantum dots. *Appl. Phys. Lett.* 108, 103108. doi:10.1063/1.4943622
- Anoop, G., Panwar, V., Kim, T. Y., and Jo, J. Y. (2017). Resistive switching in zno nanorods/graphene oxide hybrid multilayer structures. *Adv. Electron. Mat.* 3, 1600418. doi:10.1002/aelm.201600418
- Arnold, A. J., Razavieh, A., Nasr, J. R., Schulman, D. S., Eichfeld, C. M., and Das, S. (2017). Mimicking neurotransmitter release in chemical synapses via hysteresis engineering in mos2 transistors. *ACS Nano* 11, 3110–3118. doi:10.1021/acsnano.7b00113
- Bao, L., Zhu, J., Yu, Z., Jia, R., Cai, Q., Wang, Z., et al. (2019). Dual-gated mos2 neuron for neuromorphic computing. *ACS Appl. Mat. Interfaces* 11, 41482–41489. doi:10.1021/acsnano.7b00113
- Bapst, V., Keck, T., Grabska-Barwińska, A., Donner, C., Cubuk, E. D., Schoenholz, S. S., et al. (2020). Unveiling the predictive power of static structure in glassy systems. *Nat. Phys.* 16, 448–454. doi:10.1038/s41567-020-0842-8
- Bear, M., Connors, B., and Paradiso, M. (2016). *Neuroscience: Exploring the brain*. wolters kluwer, alphen aan den rijn: Netherlands.
- Bessonov, A. A., Kirikova, M. N., Petukhov, D. I., Allen, M., Ryhänen, T., and Bailey, M. J. (2015). Layered memristive and memcapacitive switches for printable electronics. *Nat. Mat.* 14, 199–204. doi:10.1038/nmat4135
- Bi, G.-q., and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi:10.1523/jneurosci.18-24-10464.1998
- Biesheuvel, P., Fu, Y., and Bazant, M. Z. (2011). Diffuse charge and faradaic reactions in porous electrodes. *Phys. Rev. E* 83, 061507. doi:10.1103/physrev.83.061507
- Borghetti, J., Snider, G. S., Kuekes, P. J., Yang, J. J., Stewart, D. R., and Williams, R. S. (2010). ‘memristive’ switches enable ‘stateful’ logic operations via material implication. *Nature* 464, 873–876. doi:10.1038/nature08940
- Buonomano, D. V., and Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi:10.1038/nrn2558
- Burr, G. W., Shelby, R. M., Sidler, S., Di Nolfo, C., Jang, J., Boybat, I., et al. (2015). Experimental demonstration and tolerancing of a large-scale neural network (165 000 synapses) using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Devices* 62, 3498–3507. doi:10.1109/ted.2015.2439635
- Cai, F., Correll, J. M., Lee, S. H., Lim, Y., Bothra, V., Zhang, Z., et al. (2019). A fully integrated reprogrammable memristor-CMOS system for efficient multiply-accumulate operations. *Nat. Electron.* 2, 290–299. doi:10.1038/s41928-019-0270-x
- Cao, Q., Tersoff, J., Farmer, D. B., Zhu, Y., and Han, S.-J. (2017). Carbon nanotube transistors scaled to a 40-nanometer footprint. *Science* 356, 1369–1372. doi:10.1126/science.aan2476

Funding

This manuscript acknowledges the funding by China Scholarship Fund (202106970006), the German Research Foundation (DFG) Projects MemDPU (Grant Nr. DU 1896/3-1), MemCrypto (Grant Nr. DU 1896/2-1) and the Open Access Publication Fund of the Thueringer Universitaets- und Landesbibliothek Jena project (433052568).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: A hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi:10.1146/annurev.neuro.31.060407.125639
- Chan, J., Venugopal, A., Pirkle, A., McDonnell, S., Hinojos, D., Magnuson, C. W., et al. (2012). Reducing extrinsic performance-limiting factors in graphene grown by chemical vapor deposition. *ACS Nano* 6, 3224–3229. doi:10.1021/nm300107f
- Chang, S., Lee, J., Chae, S., Lee, S., Liu, C., Kahng, B., et al. (2009). Occurrence of both unipolar memory and threshold resistance switching in a nio film. *Phys. Rev. Lett.* 102, 026801. doi:10.1103/physrevlett.102.026801
- Chen, P.-Y., Kadetotad, D., Xu, Z., Mohanty, A., Lin, B., Ye, J., et al. (2015). “Technology-design co-optimization of resistive cross-point array for accelerating learning algorithms on chip,” in 2015 Design, Automation & Test in Europe Conference & Exhibition (DATE) (Grenoble, France: IEEE), 854–859.
- Chen, Y., Chen, T., Xu, Z., Sun, N., and Temam, O. (2016). Dianna family: Energy-efficient hardware accelerators for machine learning. *Commun. ACM* 59, 105–112. doi:10.1145/2996864
- Chen, Y., Zhou, Y., Zhuge, F., Tian, B., Yan, M., Li, Y., et al. (2019). Graphene-ferroelectric transistors as complementary synapses for supervised learning in spiking neural network. *npj 2D Mat. Appl.* 3, 31–39. doi:10.1038/s41699-019-0114-6
- Chen, Z., Du, N., Kiani, M., Zhao, X., Skorupa, I., Schulz, S. E., et al. (2021). Second harmonic generation exploiting ultra-stable resistive switching devices for secure hardware systems. *IEEE Trans. Nanotechnol.* 21, 71–80. doi:10.1109/tnano.2021.3135713
- Cheng, P., Sun, K., and Hu, Y. H. (2016). Memristive behavior and ideal memristor of 1t phase mos2 nanosheets. *Nano Lett.* 16, 572–576. doi:10.1021/acs.nanolett.5b04260
- Cheng, X., Cheng, Z., Wang, C., Li, M., Gu, P., Yang, S., et al. (2021). Light helicity detector based on 2d magnetic semiconductor cri3. *Nat. Commun.* 12, 6874–6876. doi:10.1038/s41467-021-27218-3
- Cho, B., Song, S., Ji, Y., Kim, T.-W., and Lee, T. (2011). Organic resistive memory devices: Performance enhancement, integration, and advanced architectures. *Adv. Funct. Mat.* 21, 2806–2829. doi:10.1002/adfm.201100686
- Cho, S., Kim, S., Kim, J. H., Zhao, J., Seok, J., Keum, D. H., et al. (2015). Phase patterning for ohmic homojunction contact in mote2. *Science* 349, 625–628. doi:10.1126/science.aab3175
- Choi, H.-S., Park, Y. J., Lee, J.-H., and Kim, Y. (2019). 3-d synapse array architecture based on charge-trap flash memory for neuromorphic application. *Electronics* 9, 57. doi:10.3390/electronics9010057
- Chowdhury, A., Ayman, A., Dey, S., Sarkar, M., and Arka, A. I. (2017). “Simulations of threshold logic unit problems using memristor based synapses and cmos neuron,” in 2017 3rd International Conference on Electrical Information and Communication Technology (EICT) (Khulna, Bangladesh: IEEE), 1–4.
- Chua, L. O., and Kang, S. M. (1976). Memristive devices and systems. *Proc. IEEE* 64, 209–223. doi:10.1109/proc.1976.10092
- Chua, L. O. (2003). Nonlinear circuit foundations for nanodevices, part I: The four-element torus. *Proc. IEEE* 91, 1830–1859. doi:10.1109/jproc.2003.818319
- Danesh, C. D., Shaffer, C. M., Nathan, D., Shenoy, R., Tudor, A., Tadayon, M., et al. (2019). Synaptic resistors for concurrent inference and learning with high energy efficiency. *Adv. Mat.* 31, 1808032. doi:10.1002/adma.201808032
- Davies, M., Srinivasa, N., Lin, T.-H., China, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee Micro* 38, 82–99. doi:10.1109/mm.2018.112130359
- Deng, W., Zhang, X., Jia, R., Huang, L., Zhang, X., and Jie, J. (2019). Organic molecular crystal-based photosynaptic devices for an artificial visual-perception system. *NPG Asia Mat.* 11, 77–79. doi:10.1038/s41427-019-0182-2
- Dev, D., Krishnaprasad, A., He, Z., Das, S., Shawkat, M. S., Manley, M., et al. (2019). “Artificial neuron using ag/2d-mos2/au threshold switching memristor,” in 2019 Device Research Conference (DRC) (Chicago, IL: IEEE), 193–194.
- Dev, D., Krishnaprasad, A., Shawkat, M. S., He, Z., Das, S., Fan, D., et al. (2020). 2D MoS₂-based threshold switching memristor for artificial neuron. *IEEE Electron Device Lett.* 41, 936–939. doi:10.1109/led.2020.2988247
- Dewey, G., Chu-Kung, B., Boardman, J., Fastenau, J., Kavalieros, J., Kotlyar, R., et al. (2011). “Fabrication, characterization, and physics of iii-v heterojunction tunneling field effect transistors (h-tfet) for steep sub-threshold swing,” in 2011 International electron devices meeting (Washington, DC: IEEE), 33–36.
- Diorio, C., Hasler, P., Minch, A., and Mead, C. A. (1996). A single-transistor silicon synapse. *IEEE Trans. Electron Devices* 43, 1972–1980. doi:10.1109/16.543035
- Du, N., Schmidt, H., and Polian, I. (2021a). Low-power emerging memristive designs towards secure hardware systems for applications in internet of things. *Nano Mater. Sci.* 3, 186–204. doi:10.1016/j.nanomns.2021.01.001
- Du, N., Zhao, X., Chen, Z., Choubey, B., Di Ventra, M., Skorupa, I., et al. (2021b). Synaptic plasticity in memristive artificial synapses and their robustness against noisy inputs. *Front. Neurosci.* 696, 660894. doi:10.3389/fnins.2021.660894
- Duan, Q., Jing, Z., Zou, X., Wang, Y., Yang, K., Zhang, T., et al. (2020). Spiking neurons with spatiotemporal dynamics and gain modulation for monolithically integrated memristive neural networks. *Nat. Commun.* 11, 3399. doi:10.1038/s41467-020-17215-3
- Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., et al. (2015). “Convolutional networks on graphs for learning molecular fingerprints,” in *Advances in neural information processing systems* (London: MIT Press), 28.
- Fan, J., Xu, W., Wu, Y., and Gong, Y. (2010). Human tracking using convolutional neural networks. *IEEE Trans. Neural Netw.* 21, 1610–1623. doi:10.1109/tnn.2010.2066286
- Farris, S., and Dudek, S. M. (2015). “From where? Synaptic tagging allows the nucleus not to care,” in *Synaptic tagging and capture* (New York, NY: Springer), 143–153.
- Feng, P., Xu, W., Yang, Y., Wan, X., Shi, Y., Wan, Q., et al. (2017). Printed neuromorphic devices based on printed carbon nanotube thin-film transistors. *Adv. Funct. Mat.* 27, 1604447. doi:10.1002/adfm.201604447
- Fiori, G., Bonaccorso, F., Iannaccone, G., Palacios, T., Neumaier, D., Seabaugh, A., et al. (2014). Electronics based on two-dimensional materials. *Nat. Nanotechnol.* 9, 768–779. doi:10.1038/nnano.2014.207
- Fong, K. D., Wang, T., and Smoukov, S. K. (2017). Multidimensional performance optimization of conducting polymer-based supercapacitor electrodes. *Sustain. Energy Fuels* 1, 1857–1874. doi:10.1039/c7se00339k
- Friedmann, S., Schemmel, J., Grübl, A., Hartel, A., Hock, M., and Meier, K. (2016). Demonstrating hybrid learning in a flexible neuromorphic hardware system. *IEEE Trans. Biomed. Circuits Syst.* 11, 128–142. doi:10.1109/tbcas.2016.2579164
- Fuller, E. J., Gabaly, F. E., Léonard, F., Agarwal, S., Plimpton, S. J., Jacobs-Gedrim, R. B., et al. (2017). Li-ion synaptic transistor for low power analog computing. *Adv. Mat.* 29, 1604310. doi:10.1002/adma.201604310
- Fuller, E. J., Keene, S. T., Melianas, A., Wang, Z., Agarwal, S., Li, Y., et al. (2019). Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* 364, 570–574. doi:10.1126/science.aaw5581
- Gandhi, R., Chen, Z., Singh, N., Banerjee, K., and Lee, S. (2011). CMOS-compatible vertical-silicon-nanowire gate-all-around p-type tunneling FETs with ≤ 50 -mV/decade subthreshold swing. *IEEE Electron Device Lett.* 32, 1504–1506. doi:10.1109/led.2011.2165331
- Ganjipour, B., Wallentin, J., Borgstrom, M. T., Samuelson, L., and Thelander, C. (2012). Tunnel field-effect transistors based on in-p-gaas heterostructure nanowires. *ACS Nano* 6, 3109–3113. doi:10.1021/nm204838m
- Gao, B., Zhou, Y., Zhang, Q., Zhang, S., Yao, P., Xi, Y., et al. (2022). Memristor-based analog computing for brain-inspired sound localization with *in situ* training. *Nat. Commun.* 13, 2026–2028. doi:10.1038/s41467-022-29712-8
- Ge, J., Li, D., Huang, C., Zhao, X., Qin, J., Liu, H., et al. (2020). Memristive synapses with high reproducibility for flexible neuromorphic networks based on biological nanocomposites. *Nanoscale* 12, 720–730. doi:10.1039/c9nr08001e
- Geim, A. K., and Grigorieva, I. V. (2013). Van der waals heterostructures. *Nature* 499, 419–425. doi:10.1038/nature12385
- Gerstner, W., and Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. London, United Kingdom: Cambridge University Press.
- Go, G.-T., Lee, Y., Seo, D.-G., and Lee, T.-W. (2022). Organic neuro-electronics: From neural interface to neuroprosthetics. *Adv. Mat.*, 2201864. doi:10.1002/adma.202201864
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. London, United Kingdom: MIT press.
- Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.
- Hamilton, W., Ying, Z., and Leskovec, J. (2017). “Inductive representation learning on large graphs,” in *Advances in neural information processing systems* (Long Beach, CA: 2017-proceedings.neurips.cc), 30.
- Han, G., Lee, C., Lee, J.-E., Seo, J., Kim, M., Song, Y., et al. (2021). Alternative negative weight for simpler hardware implementation of synapse device based neuromorphic system. *Sci. Rep.* 11, 23198. doi:10.1038/s41598-021-02176-4
- Han, S.-T., Hu, L., Wang, X., Zhou, Y., Zeng, Y.-J., Ruan, S., et al. (2017). Black phosphorus quantum dots with tunable memory properties and multilevel resistive

- switching characteristics. *Adv. Sci. (Weinh)*. 4, 1600435. doi:10.1002/adv.201600435
- Hao, S., Ji, X., Zhong, S., Pang, K. Y., Lim, K. G., Chong, T. C., et al. (2020). A monolayer leaky integrate-and-fire neuron for 2d memristive neuromorphic networks. *Adv. Electron. Mat.* 6, 1901335. doi:10.1002/aelm.201901335
- Hasan, R., and Taha, T. M. (2014). "Enabling back propagation training of memristor crossbar neuromorphic processors," in 2014 International Joint Conference on Neural Networks (IJCNN) (Beijing, China: IEEE), 21–28.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.
- He, L., Liao, Z.-M., Wu, H.-C., Tian, X.-X., Xu, D.-S., Cross, G. L., et al. (2011). Memory and threshold resistance switching in ni/nio core-shell nanowires. *Nano Lett.* 11, 4601–4606. doi:10.1021/nl202017k
- Henaff, M., Bruna, J., and LeCun, Y. (2015). Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163
- Hu, M., Graves, C. E., Li, C., Li, Y., Ge, N., Montgomery, E., et al. (2018). Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mat.* 30, 1705914. doi:10.1002/adma.201705914
- Huang, Z., Wang, Z., Hu, W., Lin, C.-W., and Satoh, S. (2019). "Dot-gnn: Domain-transferred graph neural network for group re-identification," in Proceedings of the 27th ACM International Conference on Multimedia (Nice, France: Association for Computing Machinery), 1888–1896.
- Huh, W., Jang, S., Lee, J. Y., Lee, D., Lee, D., Lee, J. M., et al. (2018). 2D materials: Synaptic barristor based on phase-engineered 2D heterostructures (adv. Mater. 35/2018). *Adv. Mat.* 30, 1870266. doi:10.1002/adma.201870266
- Ielmini, D., Cagli, C., Nardi, F., and Zhang, Y. (2013). Nanowire-based resistive switching memories: Devices, operation and scaling. *J. Phys. D. Appl. Phys.* 46, 074006. doi:10.1088/0022-3727/46/7/074006
- Ilyas, A., Li, D., Li, C., Jiang, X., Jiang, Y., and Li, W. (2020). Analog switching and artificial synaptic behavior of ag/siox: Ag/tiox/p++-si memristor device. *Nanoscale Res. Lett.* 15, 30. doi:10.1186/s11671-020-3249-7
- Jackson, B. L., Rajendran, B., Corrado, G. S., Breitwisch, M., Burr, G. W., Cheek, R., et al. (2013). Nanoscale electronic synapses using phase change devices. *ACM J. Emerg. Technol. Comput. Syst.* 9, 1–20. doi:10.1145/2463585.2463588
- Jadwiszczak, J., Keane, D., Maguire, P., Cullen, C. P., Zhou, Y., Song, H., et al. (2019). Mos2 memtransistors fabricated by localized helium ion beam irradiation. *ACS Nano* 13, 14262–14273. doi:10.1021/acsnano.9b07421
- James, C. D., Aimone, J. B., Miner, N. E., Vineyard, C. M., Rothganger, F. H., Carlson, K. D., et al. (2017). A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications. *Biol. Inspired Cogn. Archit.* 19, 49–64. doi:10.1016/j.bica.2016.11.002
- Jariwala, D., Sangwan, V. K., Lauhon, L. J., Marks, T. J., and Hersam, M. C. (2014). Emerging device applications for semiconducting two-dimensional transition metal dichalcogenides. *ACS Nano* 8, 1102–1120. doi:10.1021/nn500064s
- Jeong, D. S., Kim, K. M., Kim, S., Choi, B. J., and Hwang, C. S. (2016). Memristors for energy-efficient new computing paradigms. *Adv. Electron. Mat.* 2, 1600090. doi:10.1002/aelm.201600090
- Jiang, Y., Huang, P., Zhu, D., Zhou, Z., Han, R., Liu, L., et al. (2018). Design and hardware implementation of neuromorphic systems with rram synapses and threshold-controlled neurons for pattern recognition. *IEEE Trans. Circuits Syst. I.* 65, 2726–2738. doi:10.1109/tcsi.2018.2812419
- Jin, L., Shuai, Y., Ou, X., Luo, W., Wu, C., Zhang, W., et al. (2015). Transport properties of ar+ irradiated resistive switching bifeo3 thin films. *Appl. Surf. Sci.* 336, 354–358. doi:10.1016/j.apsusc.2014.12.136
- Joshi, J., Parker, A. C., and Hsu, C.-C. (2009). "A carbon nanotube cortical neuron with spike-timing-dependent plasticity," in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE), 1651–1654.
- Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., et al. (2017). "In-datacenter performance analysis of a tensor processing unit," in Proceedings of the 44th annual international symposium on computer architecture (Toronto, ON: Association for Computing Machinery), 1–12.
- Jung, S., Lee, H., Myung, S., Kim, H., Yoon, S. K., Kwon, S.-W., et al. (2022). A crossbar array of magnetoresistive memory devices for in-memory computing. *Nature* 601, 211–216. doi:10.1038/s41586-021-04196-6
- Kalita, H., Krishnaprasad, A., Choudhary, N., Das, S., Dev, D., Ding, Y., et al. (2019). Artificial neuron using vertical mos2/graphene threshold switching memristors. *Sci. Rep.* 9, 53–58. doi:10.1038/s41598-018-35828-z
- Karpathy, A., and Fei-Fei, L. (2015). "Deep visual-semantic alignments for generating image descriptions," in Proceedings of the IEEE conference on computer vision and pattern recognition (San Juan, PR: IEEE), 3128–3137.
- Kayser, L. V., and Lipomi, D. J. (2019). Stretchable conductive polymers and composites based on pedot and pedot: Pss. *Adv. Mat.* 31, 1806133. doi:10.1002/adma.201806133
- Kendall, J. D., and Kumar, S. (2020). The building blocks of a brain-inspired computer. *Appl. Phys. Rev.* 7, 011305. doi:10.1063/1.5129306
- Kim, K., Chen, C.-L., Truong, Q., Shen, A. M., and Chen, Y. (2013). A carbon nanotube synapse with dynamic logic and learning. *Adv. Mat.* 25, 1693–1698. doi:10.1002/adma.201203116
- Kim, G., In, J. H., Kim, Y. S., Rhee, H., Park, W., Song, H., et al. (2021). Self-clocking fast and variation tolerant true random number generator based on a stochastic mott memristor. *Nat. Commun.* 12, 2906. doi:10.1038/s41467-021-23184-y
- Kim, S., Choi, B., Lim, M., Yoon, J., Lee, J., Kim, H.-D., et al. (2017). Pattern recognition using carbon nanotube synaptic transistors with an adjustable weight update protocol. *ACS Nano* 11, 2814–2822. doi:10.1021/acsnano.6b07894
- Kim, S., Du, C., Sheridan, P., Ma, W., Choi, S., and Lu, W. D. (2015a). Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* 15, 2203–2211. doi:10.1021/acs.nanolett.5b00697
- Kim, S., Myeong, G., Shin, W., Lim, H., Kim, B., Jin, T., et al. (2020). Thickness-controlled black phosphorus tunnel field-effect transistor for low-power switches. *Nat. Nanotechnol.* 15, 203–206. doi:10.1038/s41565-019-0623-7
- Kim, S., Yoon, J., Kim, H.-D., and Choi, S.-J. (2015b). Carbon nanotube synaptic transistor network for pattern recognition. *ACS Appl. Mat. Interfaces* 7, 25479–25486. doi:10.1021/acsmi.5b08541
- Knoll, L., Zhao, Q.-T., Nichau, A., Trelenkamp, S., Richter, S., Schäfer, A., et al. (2013). Inverters with strained si nanowire complementary tunnel field-effect transistors. *IEEE Electron Device Lett.* 34, 813–815. doi:10.1109/led.2013.2258652
- Ko, T.-J., Li, H., Mofid, S. A., Yoo, C., Okogbue, E., Han, S. S., et al. (2020). Two-dimensional near-atom-thickness materials for emerging neuromorphic devices and applications. *IScience* 23, 101676. doi:10.1016/j.isci.2020.101676
- Kornijcuk, V., Lim, H., Seok, J. Y., Kim, G., Kim, S. K., Kim, L., et al. (2016). Leaky integrate-and-fire neuron circuit based on floating-gate integrator. *Front. Neurosci.* 10, 212. doi:10.3389/fnins.2016.00212
- Krestinskaya, O., Ibrayev, T., and James, A. P. (2017). Hierarchical temporal memory features with memristor logic circuits for pattern recognition. *IEEE Trans. Comput.-Aided. Des. Integr. Circuits Syst.* 37, 1143–1156. doi:10.1109/tcad.2017.2748024
- Krestinskaya, O., James, A. P., and Chua, L. O. (2019). Neuromemristive circuits for edge computing: A review. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4–23. doi:10.1109/tnnls.2019.2899262
- Krishnaprasad, A., Choudhary, N., Das, S., Dev, D., Kalita, H., Chung, H.-S., et al. (2019). Electronic synapses with near-linear weight update using mos2/graphene memristors. *Appl. Phys. Lett.* 115, 103104. doi:10.1063/1.5108899
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 25.
- Kumar, S., Wang, X., Strachan, J. P., Yang, Y., and Lu, W. D. (2022). Dynamical memristors for higher-complexity neuromorphic computing. *Nat. Rev. Mat.* 7, 575–591. doi:10.1038/s41578-022-00434-z
- Kuzum, D., Yu, S., and Wong, H. P. (2013). Synaptic electronics: Materials, devices and applications. *Nanotechnology* 24, 382001. doi:10.1088/0957-4484/24/38/382001
- Kwon, K. C., Baek, J. H., Hong, K., Kim, S. Y., and Jang, H. W. (2022). Memristive devices based on two-dimensional transition metal chalcogenides for neuromorphic computing. *Nanomicro. Lett.* 14, 58–30. doi:10.1007/s40820-021-00784-3
- Lai, Q., Zhang, L., Li, Z., Stickle, W. F., Williams, R. S., and Chen, Y. (2010). Ionic/electronic hybrid materials integrated in a synaptic transistor with signal processing and learning functions. *Adv. Mat.* 22, 2448–2453. doi:10.1002/adma.201000282
- Lanza, M., Wong, H.-S. P., Pop, E., Ielmini, D., Strukov, D., Regan, B. C., et al. (2019). Recommended methods to study resistive switching devices. *Adv. Electron. Mat.* 5, 1800143. doi:10.1002/aelm.201800143
- Lavecchia, A. (2015). Machine-learning approaches in drug discovery: Methods and applications. *Drug Discov. today* 20, 318–331. doi:10.1016/j.drudis.2014.10.012
- Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* 8, 98–113. doi:10.1109/72.554195

- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* 521 (7553), 436–444. doi:10.1038/nature14539
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi:10.1109/5.726791
- Lee, D., Hwang, E., Lee, Y., Choi, Y., Kim, J. S., Lee, S., et al. (2016). Multibit mos2 photoelectronic memory with ultrahigh sensitivity. *Adv. Mat.* 28, 9196–9202. doi:10.1002/adma.201603571
- Lee, G., Baek, J.-H., Ren, F., Pearson, S. J., Lee, G.-H., and Kim, J. (2021). Artificial neuron and synapse devices based on 2d materials. *Small* 17, 2100640. doi:10.1002/sml.202100640
- Lee, H.-S., Sangwan, V. K., Rojas, W. A. G., Bergeron, H., Jeong, H. Y., Yuan, J., et al. (2020). Artificial neural networks: Dual-gated mos2 memristor crossbar array (adv. funct. mater. 45/2020). *Adv. Funct. Mater.* 30, 2070297. doi:10.1002/adfm.202070297
- Lee, K. M., Wang, D. H., Koerner, H., Vaia, R. A., Tan, L.-S., and White, T. J. (2012). Enhancement of photogenerated mechanical force in azobenzene-functionalized polyimides. *Angew. Chem. Int. Ed. Engl.* 51, 4193–4197. doi:10.1002/ange.201200726
- Lee, M.-J., Lee, C. B., Lee, D., Lee, S. R., Chang, M., and Hur, J. H. (2011). A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta2O5-x/TaO2-x bilayer structures. *Nat Mater.* 10 (8), 625–630. doi:10.1038/nmat3070
- Lee, W., Kang, S. H., Kim, J.-Y., Kolekar, G. B., Sung, Y.-E., and Han, S.-H. (2009). TiO2 nanotubes with a zno thin energy barrier for improved current efficiency of cdse quantum-dot-sensitized solar cells. *Nanotechnology* 20, 335706. doi:10.1088/0957-4484/20/33/335706
- Lee, W., and Kim, J. (2019). Accuracy investigation of a neuromorphic machine learning system due to electromagnetic noises using peec model. *IEEE Trans. Compon. Packag. Manuf. Technol.* 9, 2066–2078. doi:10.1109/tcpmt.2019.2917910
- Lei, Y., Zeng, H., Luo, W., Shuai, Y., Wei, X., Du, N., et al. (2016). Ferroelectric and flexible barrier resistive switching of epitaxial bifeo3 films studied by temperature-dependent current and capacitance spectroscopy. *J. Mat. Sci. Mat. Electron.* 27, 7927–7932. doi:10.1007/s10854-016-4784-y
- Li, B., Li, S., Wang, H., Chen, L., Liu, L., Feng, X., et al. (2020). An electronic synapse based on 2d ferroelectric cuinp2s6. *Adv. Electron. Mat.* 6, 2000760. doi:10.1002/aelm.202000760
- Li, S., Pam, M.-E., Li, Y., Chen, L., Chien, Y.-C., Fong, X., et al. (2022). Wafer-scale 2d hafnium diselenide based memristor crossbar array for energy-efficient neural network hardware. *Adv. Mater.* 34, 2103376. doi:10.1002/adma.202103376
- Li, Y., Tarlow, D., Brockschmidt, M., and Zemel, R. (2015). “Gated graph sequence neural networks,” in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016.
- Liang, K.-D., Huang, C.-H., Lai, C.-C., Huang, J.-S., Tsai, H.-W., Wang, Y.-C., et al. (2014). Single CuO_x nanowire memristor: Forming-free resistive switching behavior. *ACS Appl. Mat. Interfaces* 6, 16537–16544. doi:10.1021/am502741m
- Lin, L., Osan, R., and Tsien, J. Z. (2006). Organizing principles of real-time memory encoding: Neural clique assemblies and universal neural codes. *TRENDS Neurosci.* 29, 48–57. doi:10.1016/j.tins.2005.11.004
- Lin, Y., Wang, Z., Zhang, X., Zhang, X., Zeng, T., Bai, L., et al. (2020). Photoreduced nanocomposites of graphene oxide/N-doped carbon dots toward all-carbon memristive synapses. *NPG Asia Mater.* 12, 64. doi:10.1038/s41427-020-00245-0
- Linn, E., Rosezin, R., Kügeler, C., and Waser, R. (2010). Complementary resistive switches for passive nanocrossbar memories. *Nat. Mat.* 9, 403–406. doi:10.1038/nmat2748
- Liu, C., Chen, H., Wang, S., Liu, Q., Jiang, Y.-G., Zhang, D. W., et al. (2020). Two-dimensional materials for next-generation computing technologies. *Nat. Nanotechnol.* 15, 545–557. doi:10.1038/s41565-020-0724-3
- Liu, C., Yang, Q., Yan, B., Yang, J., Du, X., Zhu, W., et al. (2016). “A memristor crossbar based computing engine optimized for high speed and accuracy,” in 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI) (Pittsburgh, PA: IEEE), 110–115.
- Liu, X., and Zeng, Z. (2022). Memristor crossbar architectures for implementing deep neural networks. *Complex Intell. Syst.* 8, 787–802. doi:10.1007/s40747-021-00282-4
- Lu, H., and Seabaugh, A. (2014). Tunnel field-effect transistors: State-of-the-art. *IEEE J. Electron Devices Soc.* 2, 44–49. doi:10.1109/jeds.2014.2326622
- Luong, M.-T., Pham, H., and Manning, C. D. (2015). “Effective approaches to attention-based neural machine translation,” in EMNLP 2015 tenth workshop on statistical machine translation, Lisbon, Portugal, September 17–18, 2015
- Ma, T., Chen, J., and Xiao, C. (2018). “Constrained generation of semantically valid graphs via regularizing variational autoencoders,” in *Advances in neural information processing systems* (Montréal, QC: Curran Associates Inc), 31.
- Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Netw.* 10, 1659–1671. doi:10.1016/s0893-6080(97)00011-7
- Maier, P., Hartmann, F., Emmerling, M., Schneider, C., Kamp, M., Höfling, S., et al. (2016). Electro-photo-sensitive memristor for neuromorphic and arithmetic computing. *Phys. Rev. Appl.* 5, 054011. doi:10.1103/physrevapplied.5.054011
- Maier, P., Hartmann, F., Mauder, T., Emmerling, M., Schneider, C., Kamp, M., et al. (2015). Memristive operation mode of a site-controlled quantum dot floating gate transistor. *Appl. Phys. Lett.* 106, 203501. doi:10.1063/1.4921061
- McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133. doi:10.1007/bf02478259
- Meier, D., and Selbach, S. M. (2022). Ferroelectric domain walls for nanotechnology. *Nat. Rev. Mat.* 7, 157–173. doi:10.1038/s41578-021-00375-z
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 668–673. doi:10.1126/science.1254642
- Mesaritakis, C., Kapsalis, A., Bogris, A., and Syvridis, D. (2016). Artificial neuron based on integrated semiconductor quantum dot mode-locked lasers. *Sci. Rep.* 6, 39317. doi:10.1038/srep39317
- Milano, G., Luebben, M., Ma, Z., Dunin-Borkowski, R., Boarino, L., Pirri, C. F., et al. (2018a). Self-limited single nanowire systems combining all-in-one memristive and neuromorphic functionalities. *Nat. Commun.* 9, 5151. doi:10.1038/s41467-018-07330-7
- Milano, G., Porro, S., Ali, M. Y., Bejtka, K., Bianco, S., Beccaria, F., et al. (2018b). Unravelling resistive switching mechanism in zno nw arrays: The role of the polycrystalline base layer. *J. Phys. Chem. C* 122, 866–874. doi:10.1021/acs.jpcc.7b09978
- Mokerov, V., Fedorov, Y. V., Velikovski, L., and Scherbakova, M. Y. (2001). New quantum dot transistor. *Nanotechnology* 12, 552–555. doi:10.1088/0957-4484/12/4/336
- Moon, J., Ma, W., Shin, J. H., Cai, F., Du, C., Lee, S. H., et al. (2019). Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat. Electron.* 2, 480–487. doi:10.1038/s41928-019-0313-3
- Moreels, I., Lambert, K., De Muynck, D., Vanhaecke, F., Poelman, D., Martins, J. C., et al. (2007). Composition and size-dependent extinction coefficient of colloidal pbse quantum dots. *Chem. Mat.* 19, 6101–6106. doi:10.1021/cm071410q
- Mouttet, B. (2010). “Memristive systems analysis of 3-terminal devices,” in 2010 17th IEEE International Conference on Electronics, Circuits and Systems (Athens, Greece: IEEE), 930–933.
- Mukherjee, S., Maiti, R., Katiyar, A. K., Das, S., and Ray, S. K. (2016). Novel colloidal mos2 quantum dot heterojunctions on silicon platforms for multifunctional optoelectronic devices. *Sci. Rep.* 6, 29016. doi:10.1038/srep29016
- Nagashima, K., Yanagida, T., Oka, K., Kanai, M., Klamchuen, A., Kim, J.-S., et al. (2011). Intrinsic mechanisms of memristive switching. *Nano Lett.* 11, 2114–2118. doi:10.1021/nl200707n
- Nagashima, K., Yanagida, T., Oka, K., Taniguchi, M., Kawai, T., Kim, J.-S., et al. (2010). Resistive switching multistate nonvolatile memory effects in a single cobalt oxide nanowire. *Nano Lett.* 10, 1359–1363. doi:10.1021/nl9042906
- Namsheer, K., and Rout, C. (2021). Conducting polymers: A comprehensive review on recent advances in synthesis, properties and applications. *rsc Adv.* 11, 5659–5697. doi:10.1039/d0ra07800j
- Neckar, A., Fok, S., Benjamin, B. V., Stewart, T. C., Oza, N. N., Voelker, A. R., et al. (2018). Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model. *Proc. IEEE* 107, 144–164. doi:10.1109/jproc.2018.2881432
- Niepert, M., Ahmed, M., and Kutzkov, K. (2016). “Learning convolutional neural networks for graphs,” in International conference on machine learning (PMLR), 2014–2023.
- Nikam, H., Satyam, S., and Sahay, S. (2021). Long short-term memory implementation exploiting passive rram crossbar array. *IEEE Trans. Electron Devices* 69, 1743–1751. doi:10.1109/ted.2021.3133197
- Nowshin, F., and Yi, Y. (2022). “Memristor-based deep spiking neural network with a computing-in-memory architecture,” in 23rd International Symposium on Quality Electronic Design (ISQED), Santa Clara, CA, April 6–7, 2022, 1–6. doi:10.1109/ISQED54688.2022.9806206
- Oka, K., Yanagida, T., Nagashima, K., Tanaka, H., and Kawai, T. (2009). Nonvolatile bipolar resistive memory switching in single crystalline nio

- heterostructured nanowires. *J. Am. Chem. Soc.* 131, 3434–3435. doi:10.1021/ja8089922
- O'Kelly, C., Fairfield, J. A., and Boland, J. J. (2014). A single nanoscale junction with programmable multilevel memory. *ACS Nano* 8, 11724–11729. doi:10.1021/nn505139m
- Paimkras, E., Plana, L. A., Garside, J., Temple, S., Galluppi, F., Patterson, C., et al. (2013). Spinnaker: A 1-w 18-core system-on-chip for massively-parallel neural network simulation. *IEEE J. Solid-State Circuits* 48, 1943–1953. doi:10.1109/jssc.2013.2259038
- Park, J., Lee, S., Lee, J., and Yong, K. (2013). A light incident angle switchable zno nanorod memristor: Reversible switching behavior between two non-volatile memory devices. *Adv. Mat.* 25, 6423–6429. doi:10.1002/adma.201303017
- Pei, J., Deng, L., Song, S., Zhao, M., Zhang, Y., Wu, S., et al. (2019). Towards artificial general intelligence with hybrid tianjin chip architecture. *Nature* 572, 106–111. doi:10.1038/s41586-019-1424-8
- Pershin, Y. V., and Di Ventra, M. (2011). Memory effects in complex materials and nanoscale systems. *Adv. Phys.* 60, 145–227. doi:10.1080/00018732.2010.544961
- Pi, S., Li, C., Jiang, H., Xia, W., Xin, H., Yang, J. J., et al. (2019). Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension. *Nat. Nanotechnol.* 14, 35–39. doi:10.1038/s41565-018-0302-0
- Pickett, M. D., and Williams, R. S. (2012). Sub-100 fJ and sub-nanosecond thermally driven threshold switching in niobium oxide crosspoint nanodevices. *Nanotechnology* 23, 215202. doi:10.1088/0957-4484/23/21/215202
- Porro, S., Risplendi, F., Cicero, G., Bejtka, K., Milano, G., Rivolo, P., et al. (2017). Multiple resistive switching in core-shell zno nanowires exhibiting tunable surface states. *J. Mat. Chem. C Mat.* 5, 10517–10523. doi:10.1039/c7tc02383a
- Pyarasani, R. D., Jayaramudu, T., and John, A. (2019). Polyaniline-based conducting hydrogels. *J. Mat. Sci.* 54, 974–996. doi:10.1007/s10853-018-2977-x
- Qi, J., Huang, J., Paul, D., Ren, J., Chu, S., and Liu, J. (2013). Current self-compliance and self-rectifying resistive switching in ag-electroded single na-doped zno nanowires. *Nanoscale* 5, 2651–2654. doi:10.1039/c3nr00027c
- Qian, C., Kong, L.-a., Yang, J., Gao, Y., and Sun, J. (2017). Multi-gate organic neuron transistors for spatiotemporal information processing. *Appl. Phys. Lett.* 110, 083302. doi:10.1063/1.4977069
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems* (Montreal, QC: MIT Press), 28.
- Rivnay, J., Inal, S., Salleo, A., Owens, R. M., Berggren, M., and Malliaras, G. G. (2018). Organic electrochemical transistors. *Nat. Rev. Mat.* 3, 17086. doi:10.1038/natrevmats.2017.86
- Roy, K., Jaiswal, A., and Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575, 607–617. doi:10.1038/s41586-019-1677-2
- Sanchez Esqueda, I., Yan, X., Rutherglen, C., Kane, A., Cain, T., Marsh, P., et al. (2018). Aligned carbon nanotube synaptic transistors for large-scale neuromorphic computing. *ACS Nano* 12, 7352–7361. doi:10.1021/acsnano.8b03831
- Sangwan, V. K., and Hersam, M. C. (2018). Electronic transport in two-dimensional materials. arXiv preprint arXiv:1802.01045.
- Sangwan, V. K., and Hersam, M. C. (2020). Neuromorphic nanoelectronic materials. *Nat. Nanotechnol.* 15, 517–528. doi:10.1038/s41565-020-0647-z
- Sangwan, V. K., Jariwala, D., Kim, I. S., Chen, K.-S., Marks, T. J., Lauhon, L. J., et al. (2015). Gate-tunable memristive phenomena mediated by grain boundaries in single-layer mos₂. *Nat. Nanotechnol.* 10, 403–406. doi:10.1038/nnano.2015.56
- Sangwan, V. K., Lee, H.-S., Bergeron, H., Balla, I., Beck, M. E., Chen, K.-S., et al. (2018). Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide. *Nature* 554, 500–504. doi:10.1038/nature25747
- Schmidt, H., Shuai, Y., Zhou, S., Skorupa, I., Ou, X., Du, N., et al. (2016). “Integrated non-volatile memory elements, design and use,” in *US patent* (Dresden, Germany: US Patent Office), 9, 520 445.
- Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., et al. (2017). A survey of neuromorphic computing and neural networks in hardware. arXiv preprint arXiv:1705.06963
- Schwerdt, H. N., Zhang, E., Kim, M. J., Yoshida, T., Stanwicks, L., Amemori, S., et al. (2018). Cellular-scale probes enable stable chronic subsecond monitoring of dopamine neurochemicals in a rodent model. *Commun. Biol.* 1, 144. doi:10.1038/s42003-018-0147-y
- Seo, S., Jo, S.-H., Kim, S., Shim, J., Oh, S., Kim, J.-H., et al. (2018). Artificial optical synapse for colored and color-mixed pattern recognition. *Nat. Commun.* 9 (1), 5106. doi:10.1038/s41467-018-07572-5
- Seok, J. Y., Song, S. J., Yoon, J. H., Yoon, K. J., Park, T. H., Kwon, D. E., et al. (2014). A review of three-dimensional resistive switching cross-bar array memories from the integration and materials property points of view. *Adv. Funct. Mat.* 24, 5316–5339. doi:10.1002/adfm.201303520
- Shen, A. M., Chen, C.-L., Kim, K., Cho, B., Tudor, A., and Chen, Y. (2013). Analog neuromorphic module based on carbon nanotube synapses. *ACS Nano* 7, 6117–6122. doi:10.1021/nn401946s
- Sheridan, P. M., Cai, F., Du, C., Ma, W., Zhang, Z., and Lu, W. D. (2017). Sparse coding with memristor networks. *Nat. Nanotechnol.* 12, 784–789. doi:10.1038/nnano.2017.83
- Shi, Y., Liang, X., Yuan, B., Chen, V., Li, H., Hui, F., et al. (2018). Electronic synapses made of layered two-dimensional materials. *Nat. Electron.* 1, 458–465. doi:10.1038/s41928-018-0118-9
- Shi, Y., Pan, L., Liu, B., Wang, Y., Cui, Y., Bao, Z., et al. (2014). Nanostructured conductive polypyrrole hydrogels as high-performance, flexible supercapacitor electrodes. *J. Mat. Chem. A* 2, 6086–6091. doi:10.1039/c4ta00484a
- Shin, T., Park, S., Kim, S., Park, H., Lho, D., Kim, S., et al. (2020). “Modeling and demonstration of hardware-based deep neural network (dnn) inference using memristor crossbar array considering signal integrity,” in 2020 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI) (Reno, NV: IEEE), 417–421.
- Simonyan, K., and Zisserman, A. (2014). “Very deep convolutional networks for large-scale image recognition,” in EMNLP 2015 tenth workshop on statistical machine translation, Lisbon, Portugal, September 17–18, 2015.
- Sipos, B., Kusmartseva, A. F., Akrap, A., Berger, H., Forró, L., and Tutiš, E. (2008). From mott state to superconductivity in 1t-tas₂. *Nat. Mat.* 7, 960–965. doi:10.1038/nmat2318
- Sivaramakrishnan, S., Sterbing-D'Angelo, S. J., Filipovic, B., D'Angelo, W. R., Oliver, D. L., and Kuwada, S. (2004). Gabaa synapses shape neuronal responses to sound intensity in the inferior colliculus. *J. Neurosci.* 24, 5031–5043. doi:10.1523/jneurosci.0357-04.2004
- Smagulova, K., Krestinskaya, O., and James, A. P. (2018). A memristor-based long short term memory circuit. *Analog Integr. Circ. Sig. Process* 95, 467–472. doi:10.1007/s10470-018-1180-y
- Smagulova, K., and James, A. P. (2019). A survey on LSTM memristive neural network architectures and applications. *Eur. Phys. J. Spec. Top.* 228, 2313–2324. doi:10.1140/epjst/e2019-900046-x
- Stetler, R. A., Gao, Y., Signore, A., Cao, G., and Chen, J. (2009). Hsp27: Mechanisms of cellular protection against neuronal injury. *Curr. Mol. Med.* 9, 863–872. doi:10.2174/156652409789105561
- Stojchevska, L., Vaskivskiy, I., Mertelj, T., Kusar, P., Svetin, D., Brazovskii, S., et al. (2014). Ultrafast switching to a stable hidden quantum state in an electronic crystal. *Science* 344, 177–180. doi:10.1126/science.1241591
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Thirty-first AAAI conference on artificial intelligence (San Francisco, CA: AAAI Press).
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition (Boston, MA: IEEE), 1–9.
- Tan, C., Liu, Z., Huang, W., and Zhang, H. (2015). Non-volatile resistive memory devices based on solution-processed ultrathin two-dimensional nanomaterials. *Chem. Soc. Rev.* 44, 2615–2628. doi:10.1039/c4cs00399c
- Tang, J., Yuan, F., Shen, X., Wang, Z., Rao, M., He, Y., et al. (2019). Bridging biological and artificial neural networks with emerging neuromorphic devices: Fundamentals, progress, and challenges. *Adv. Mat.* 31, 1902761. doi:10.1002/adma.201902761
- Tang, Z., Zhu, R., Hu, R., Chen, Y., Wu, E. Q., Wang, H., et al. (2020). A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors. *IEEE Trans. Cogn. Dev. Syst.* 13, 645–656. doi:10.1109/tcds.2020.3003377
- Thomas, A., Resmi, A., Ganguly, A., and Jinesh, K. (2020). Programmable electronic synapse and nonvolatile resistive switches using mos₂ quantum dots. *Sci. Rep.* 10, 12450. doi:10.1038/s41598-020-68822-5
- Tian, H., Zhao, L., Wang, X., Yeh, Y.-W., Yao, N., Rand, B. P., et al. (2017). Extremely low operating current resistive memory based on exfoliated 2d perovskite single crystals for neuromorphic computing. *ACS Nano* 11, 12247–12256. doi:10.1021/acsnano.7b05726
- Tornig, W., and Altman, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inf. Model.* 59, 4131–4149. doi:10.1021/acs.jcim.9b00628

- Tuchman, Y., Mangoma, T. N., Gkoupidenis, P., Van De Burgt, Y., John, R. A., Mathews, N., et al. (2020). Organic neuromorphic devices: Past, present, and future challenges. *MRS Bull.* 45, 619–630. doi:10.1557/mrs.2020.196
- Valov, I., Waser, R., Jameson, J. R., and Kozicki, M. N. (2011). Electrochemical metallization memories—Fundamentals, applications, prospects. *Nanotechnology* 22, 289502. doi:10.1088/0957-4484/22/28/289502
- Van De Burgt, Y., Lubberman, E., Fuller, E. J., Keene, S. T., Faria, G. C., Agarwal, S., et al. (2017). A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mat.* 16, 414–418. doi:10.1038/nmat4856
- van De Burgt, Y., Melianas, A., Keene, S. T., Malliaras, G., and Salleo, A. (2018). Organic electronics for neuromorphic computing. *Nat. Electron.* 1, 386–397. doi:10.1038/s41928-018-0103-3
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). “Graph attention networks,” in The 6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, April 30–May 3, 2018.
- Vu, Q. A., Shin, Y. S., Kim, Y. R., Nguyen, V. L., Kang, W. T., Kim, H., et al. (2016). Two-terminal floating-gate memory with van der waals heterostructures for ultrahigh on/off ratio. *Nat. Commun.* 7, 12725–12728. doi:10.1038/ncomms12725
- Wang, D. H., Lee, K. M., Yu, Z., Koerner, H., Vaia, R. A., White, T. J., et al. (2011). Photochemical response of glassy azobenzene polyimide networks. *Macromolecules* 44, 3840–3846. doi:10.1021/ma200427q
- Wang, H., Lu, W., Hou, S., Yu, B., Zhou, Z., Xue, Y., et al. (2020). A 2d-sns film with ferroelectricity and its bio-realistic synapse application. *Nanoscale* 12, 21913–21922. doi:10.1039/d0nr03724a
- Wang, L., Liao, W., Wong, S. L., Yu, Z. G., Li, S., Lim, Y.-F., et al. (2019). Artificial synapses based on multiterminal memtransistors for neuromorphic application. *Adv. Funct. Mat.* 29, 1901106. doi:10.1002/adfm.201901106
- Wang, S., Chen, C., Yu, Z., He, Y., Chen, X., Wan, Q., et al. (2019a). A mos2/ptcd hybrid heterojunction synapse with efficient photoelectric dual modulation and versatility. *Adv. Mat.* 31, 1806227. doi:10.1002/adma.201806227
- Wang, S., Hou, X., Liu, L., Li, J., Shan, Y., Wu, S., et al. (2019b). A photoelectric-stimulated mos2 transistor for neuromorphic engineering. *Research* 2019, 10. doi:10.34133/2019/1618798
- Wang, T.-Y., Meng, J.-L., He, Z.-Y., Chen, L., Zhu, H., Sun, Q.-Q., et al. (2020). Ultralow power wearable heterosynapse with photoelectric synergistic modulation. *Adv. Sci. (Weinh.)* 7, 1903480. doi:10.1002/advs.201903480
- Wang, Z., Joshi, S., Savelev, S. E., Jiang, H., Midya, R., Lin, P., et al. (2017). Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mat.* 16, 101–108. doi:10.1038/nmat4756
- Wang, Z., Joshi, S., Savelev, S., Song, W., Midya, R., Li, Y., et al. (2018). Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* 1, 137–145. doi:10.1038/s41928-018-0023-2
- Waser, R., Dittmann, R., Staikov, G., and Szot, K. (2009). Redox-based resistive switching memories—nanoionic mechanisms, prospects, and challenges. *Adv. Mat.* 21, 2632–2663. doi:10.1002/adma.200900375
- Weston, J., Chopra, S., and Bordes, A. (2014). “Memory networks,” in EMNLP 2015 tenth workshop on statistical machine translation, Lisbon, Portugal, September 17–18, 2015
- Wu, J., Yuan, H., Meng, M., Chen, C., Sun, Y., Chen, Z., et al. (2017). High electron mobility and quantum oscillations in non-encapsulated ultrathin semiconducting b2o2se. *Nat. Nanotechnol.* 12, 530–534. doi:10.1038/nnano.2017.43
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., et al. (2016). “Google’s neural machine translation system: Bridging the gap between human and machine translation” in EMNLP 2017 : Conference on Empirical Methods in Natural Language Processing, Copenhagen, September 7–11, 2017
- Xia, Q., Pickett, M. D., Yang, J. J., Li, X., Wu, W., Medeiros-Ribeiro, G., et al. (2011). Two- and three-terminal resistive switches: Nanometer-scale memristors and memistors. *Adv. Funct. Mat.* 21, 2660–2665. doi:10.1002/adfm.201100180
- Xia, Q., and Yang, J. J. (2019). Memristive crossbar arrays for brain-inspired computing. *Nat. Mat.* 18, 309–323. doi:10.1038/s41563-019-0291-x
- Xiao, M., Musselman, K. P., Duley, W. W., and Zhou, Y. N. (2017). Reliable and low-power multilevel resistive switching in TiO₂ nanorod arrays structured with a TiO_x seed layer. *ACS Appl. Mat. Interfaces* 9, 4808–4817. doi:10.1021/acsami.6b14206
- Xiong, F., Liao, A. D., Estrada, D., and Pop, E. (2011). Low-power switching of phase-change materials with carbon nanotube electrodes. *Science* 332, 568–570. doi:10.1126/science.1201938
- Xu, B., Pei, J., Feng, L., and Zhang, X.-D. (2021). Graphene and graphene-related materials as brain electrodes. *J. Mater. Chem. B* 9, 9485–9496. doi:10.1039/D1TB01795K
- Xu, R., Jang, H., Lee, M.-H., Amanov, D., Cho, Y., Kim, H., et al. (2019). Vertical MoS₂ double-layer memristor with electrochemical metallization as an atomic-scale synapse with switching thresholds approaching 100 mV. *Nano. Lett.* 19 (4), 2411–2417. doi:10.1021/acs.nanolett.8b05140
- Yakopcic, C., Alom, M. Z., and Taha, T. M. (2017). “Extremely parallel memristor crossbar architecture for convolutional neural network implementation,” in 2017 International Joint Conference on Neural Networks (IJCNN) (Vancouver, BC: IEEE), 1696–1703.
- Yakopcic, C., Alom, M. Z., and Taha, T. M. (2016). “Memristor crossbar deep network implementation based on a convolutional neural network,” in 2016 International joint conference on neural networks (IJCNN) (Anchorage, Alaska: IEEE), 963–970.
- Yan, M., Deng, L., Hu, X., Liang, L., Feng, Y., Ye, X., et al. (2020). “Hygcn: A gcn accelerator with hybrid architecture,” in 2020 IEEE International Symposium on High Performance Computer Architecture (HPCA) (San Diego, CA: IEEE), 15–29.
- Yan, X., Qin, C., Lu, C., Zhao, J., Zhao, R., Ren, D., et al. (2019). Robust Ag/ZrO₂/WS₂/Pt memristor for neuromorphic computing. *ACS Appl. Mater. Interfaces* 11 (51), 48029–48038. doi:10.1021/acsami.9b17160
- Yan, X., Zhao, Q., Chen, A. P., Zhao, J., Zhou, Z., Wang, J., et al. (2019). Vacancy-induced synaptic behavior in 2D WS₂ nanosheet-based memristor for low-power neuromorphic computing. *Small* 15 (24), e1901423. doi:10.1002/smll.201901423
- Yang, J.-Q., Wang, R., Wang, Z.-P., Ma, Q.-Y., Mao, J.-Y., Ren, Y., et al. (2020). Leaky integrate-and-fire neurons based on perovskite memristor for spiking neural networks. *Nano Energy* 74, 104828. doi:10.1016/j.nanoen.2020.104828
- Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A., Stewart, D. R., and Williams, R. S. (2008). Memristive switching mechanism for metal/oxide/metal nanodevices. *Nat. Nanotechnol.* 3, 429–433. doi:10.1038/nnano.2008.160
- Yang, J., Strukov, D., and Stewart, D. (2013). Memristive devices for computing. *Nat. Nanotech.* 8 (1), 13–24. doi:10.1038/nnano.2012.240
- Yang, X., Taylor, B., Wu, A., Chen, Y., and Chua, L. O. (2022). Research progress on memristor: From synapses to computing systems. *IEEE Trans. Circuits Syst. I* 69, 1845–1857. doi:10.1109/tcsi.2022.3159153
- Yang, Y., Du, H., Xue, Q., Wei, X., Yang, Z., Xu, C., et al. (2019). Three-terminal memtransistors based on two-dimensional layered gallium selenide nanosheets for potential low-power electronics applications. *Nano Energy* 57, 566–573. doi:10.1016/j.nanoen.2018.12.057
- Yang, Y., Zhang, X., Gao, M., Zeng, F., Zhou, W., Xie, S., et al. (2011). Nonvolatile resistive switching in single crystalline zno nanowires. *Nanoscale* 3, 1917–1921. doi:10.1039/c1nr10096c
- Yao, P., Wu, H., Gao, B., Eryilmaz, S. B., Huang, X., Zhang, W., et al. (2017). Face classification using electronic synapses. *Nat. Commun.* 8, 15199. doi:10.1038/ncomms15199
- Yao, P., Wu, H., Gao, B., Tang, J., Zhang, Q., Zhang, W., et al. (2020). Fully hardware-implemented memristor convolutional neural network. *Nature* 577, 641–646. doi:10.1038/s41586-020-1942-4
- Yin, S., Song, C., Sun, Y., Qiao, L., Wang, B., Sun, Y., et al. (2019). Electric and light dual-gate tunable mos2 memristor. *ACS Appl. Mat. Interfaces* 11, 43344–43350. doi:10.1021/acsami.9b14259
- Yoshida, M., Suzuki, R., Zhang, Y., Nakano, M., and Iwasa, Y. (2015). Memristive phase switching in two-dimensional 1t-tas2 crystals. *Sci. Adv.* 1, e1500606. doi:10.1126/sciadv.1500606
- Younis, A., Chu, D., Lin, X., Yi, J., Dang, F., and Li, S. (2013). High-performance nanocomposite based memristor with controlled quantum dots as charge traps. *ACS Appl. Mat. Interfaces* 5, 2249–2254. doi:10.1021/am400168m
- Yu, S., and Chen, P.-Y. (2016). Emerging memory technologies: Recent trends and prospects. *IEEE Solid-State Circuits Mag.* 8, 43–56. doi:10.1109/mssc.2016.2546199
- Yu, S., Wu, Y., Jeyasingh, R., Kuzum, D., and Wong, H.-S. P. (2011). An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. *IEEE Trans. Electron Devices* 58, 2729–2737. doi:10.1109/ted.2011.2147791
- Yujia, L., Daniel, T., Marc, B., and Richard, Z. (2016). “Gated graph sequence neural networks,” in International Conference on Learning Representations (San Juan, Puerto Rico: DBIP).
- Zestos, A. G., Jacobs, C. B., Trikantopoulos, E., Ross, A. E., and Venton, B. J. (2014). Polyethyleneimine carbon nanotube fiber electrodes for enhanced detection of neurotransmitters. *Anal. Chem.* 86, 8568–8575. doi:10.1021/ac5003273
- Zhang, B., Shi, L., and Song, S. (2016). Creating more intelligent robots through brain-inspired computing. *Sci. Robotics* 354, 1445. doi:10.1126/science.354.6318.1445-b
- Zhang, F., and Hu, M. (2018). Memristor-based deep convolution neural network: A case study. arXiv preprint arXiv:1810.02225

- Zhang, F., Zhang, H., Krylyuk, S., Milligan, C. A., Zhu, Y., Zemlyanov, D. Y., et al. (2019). Electric-field induced structural transition in vertical mote2-and mo1-xwte2-based resistive memories. *Nat. Mat.* 18, 55–61. doi:10.1038/s41563-018-0234-y
- Zhang, X., Wang, W., Liu, Q., Zhao, X., Wei, J., Cao, R., et al. (2017). An artificial neuron based on a threshold switching memristor. *IEEE Electron Device Lett.* 39, 308–311. doi:10.1109/led.2017.2782752
- Zhang, X., Zhuo, Y., Luo, Q., Wu, Z., Midya, R., Wang, Z., et al. (2020). An artificial spiking afferent nerve based on mott memristors for neurorobotics. *Nat. Commun.* 11, 51–59. doi:10.1038/s41467-019-13827-6
- Zhang, Y., Li, Y., Wang, X., and Friedman, E. G. (2017). Synaptic characteristics of ag/aginsbte/ta-based memristor for pattern recognition applications. *IEEE Trans. Electron Devices* 64, 1806–1811. doi:10.1109/ted.2017.2671433
- Zhao, H., Tu, H., Wei, F., and Du, J. (2014). "Highly transparent dysprosium oxide-based RRAM with multilayer graphene electrode for low-power nonvolatile memory application," in *IEEE Transactions on Electron Devices*, 61, 1388–1393. doi:10.1109/TED.2014.2312611
- Zhitenev, N., Sidorenko, A., Tennant, D., and Cirelli, R. (2007). Chemical modification of the electronic conducting states in polymer nanodevices. *Nat. Nanotechnol.* 2, 237–242. doi:10.1038/nnano.2007.75
- Zhong, Y.-N., Gao, X., Xu, J.-L., Sirringhaus, H., and Wang, S.-D. (2020). Selective uv-gating organic memtransistors with modulable levels of synaptic plasticity. *Adv. Electron. Mat.* 6, 1900955. doi:10.1002/aelm.201900955
- Zhong, Y., Tang, J., Li, X., Gao, B., Qian, H., and Wu, H. (2021). Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing. *Nat. Commun.* 12, 408. doi:10.1038/s41467-020-20692-1
- Zhu, X., Li, D., Liang, X., and Lu, W. D. (2019). Ionic modulation and ionic coupling effects in mos2 devices for neuromorphic computing. *Nat. Mat.* 18, 141–148. doi:10.1038/s41563-018-0248-5
- Zhu, J., Yang, Y., Jia, R., Liang, Z., Zhu, W., Rehman, Z. U., et al. (2019). Ion gated synaptic transistors based on 2D van der Waals crystals with tunable diffusive dynamics. *Adv. Mater.* 30 (21), e1800195. doi:10.1002/adma.201800195